A Work Project presented as part of the requirements for the Award of a Master's degree in Economics from the Nova School of Business and Economics.

BEHAVIOURAL INSIGHTS ON COVID-19 IN PORTUGAL: AN ANALYSIS OF THE DETERMINANTS OF THE PROBABILITY OF PERCEIVING A HIGH RISK REGARDING THE NEW CORONAVIRUS

DIOGO VIEIRA NUNES DA CONCEIÇÃO

Work project carried out under the supervision of:

Professor Pedro Pita Barros

14-01-2021

Abstract

This work project attempts to understand what determines risk perceptions regarding COVID-19, at an individual, familiar and communitarian levels, to better design effective health safety policies and advices. It uses data collected by the company GFK in Portugal between the 23rd of May and the 30th of June 2020. It does so through the application of LPM, MLE and SUR models. As lower risk perceptions could motivate targeted interventions, the fact that young age and low levels of education are associated with such perceptions makes them natural groups for these policies in order to better control the spread of COVID-19.

Keywords: COVID-19; Risk Perceptions; Health Economics; Behavioural Economics.

Acknowledgments: I would like to thank Professor Pedro Pita Barros for his great advice and help throughout the whole process of developing this work project and for providing me with the necessary data. I would also like to thank Professor Alexander Coutts for reviewing all the material I sent to him throughout the semester and for giving me his crucial opinion on the topics of this master's thesis.

1. Introduction

On December 31st, 2019,[1] the country office of the World Health Organization (WHO) in China picked up, for the first time, on official information regarding a cluster of "viral pneumonia" in the Wuhan region. At this point in time, nobody knew what it was, and most countries kept on with "business as usual" while waiting to see how the situation would unfold. As the number of cases began to rise, so did the number of names in the list of countries that started registering cases of the new Coronavirus. Countries began to fear the nefarious consequences of what would later be declared a pandemic crisis by the WHO.

It was this fear that created the need for a research paper establishing *what determines risk perceptions regarding COVID-19 at three different proximity levels to a given person*, departing from existing notions about individual profiles and associated risk notions to compare them with actual empirical evidence. I.e. it starts with the idea that a higher positive change in the number of new COVID-19 cases, reporting to have chronical diseases related to the lungs or heart, not having a private health insurance, smoking, drinking, reporting a notion of a generally bad health state, being relatively older and living with a large number of people are all associated with being more likely to perceive a high risk-level regarding the virus and compares it with data. In general terms, this study confirms such relationships regarding the evolution of the number of cases, smoking, drinking, age and general health state notion, while contradicting that concerning the number of household member, among other revealing results.

With the world trembling in fear as the unknown future glared down at it, the chance of a global recession, the loss of millions of lives, and a disrupting force so great that threatened to change everyone's routines forever was enough to drive some to panic. The global level of uncertainty inherent to a new virus led some people to fear what they did not understand and try to find

[1] All information regarding the spreading of the new Coronavirus for an introductory contextualization was based on the World Health Organization's COVID-19 response timeline.

patterns by themselves, in order to better prepare for it. While specialists attempted to identify who was most vulnerable to the new Coronavirus, finding compelling evidence that it was most harmful for the elderly rather than the youth, some people started to establish a connection between this reduced empirical evidence and the fact that cases continued to surge. This led some to sustain the idea that, as younger people were less likely to suffer greatly from the virus, they could be asymptomatic and unknowingly spreading the virus as they went about their lives, thus harming those that were said to be most vulnerable. The truth is that, as much as this line of thought could make sense for some at the time, this unprecedented pandemic crisis made it extremely difficult to have enough data to have strong statistical proof of any such connection.

This makes communication crucial in dealing with the pandemic crisis. In a world of uncertainty, it is pivotal that those in charge maintain a clear and constant stream of information to the public. It is normal that, in a completely new situation, in which nobody really knows how to make the world go back to how it was previously, not even governments or the WHO have every single solution to all problems constantly surging. However, if they want societies to listen to them and take the necessary precautions to stop the spread of the Coronavirus, they need to have a strong communication strategy so as to influence the human perception of risk related to the pandemic and thus, contribute to the effectiveness of any measures they intend to put in place. Risk perceptions have a role to play on how people react to certain stimuli.

The present research paper helps orient communication policies in these times of crisis. It identifies patterns in risk perception that allow health authorities to target their interventions towards groups that may be more likely to underestimate the risk associated with the new Coronavirus. Thus, it starts by performing a Literature Review of several papers on the pandemic crisis, released in the year 2020, and on behavioural theories related to risk perception, stimuli and decision making. After presenting the survey that provided the Data for its econometric analysis, it moves on to explaining its Methodology and Results, finalizing with the main Conclusions, which can be useful for anyone trying to better understand what determines one's risk perception on the context of the new Coronavirus pandemic, and some possible Limitations of the present study.

2. Literature Review

During the first months of the current pandemic crisis, a large number of researchers attempted not only to take advantage of such a rare situation to further their work with extremely rich data, but also to do whatever they could to provide the public and governmental institutions with any information on the virus that could prove useful in defeating this invisible enemy.

In an early attempt to understand how adults in the United States (US) perceived the then novel Coronavirus outbreak, McFadden et al. (2020) submitted a paper in February that presented the results of a survey to the adult US population. Results showed that most people in their sample became aware of the crisis through the news and, most importantly, that, when asking ten questions related to the interviewees' risk perception in which they had to rate it in a 5-point Likert Scale,[2] the mean risk perception score, for the whole 10 questions, was 5.0 out of 10, with a slightly higher weight on values closer to the higher end. At this point in time, most also seemed to favour strict preventive measures, such as traveling restrictions (75% out of 519 individuals) and quarantining (83% out of 571 individuals).

Another paper trying to uncover the factors that could influence one's perception of risk concerning the new Coronavirus delved into the age differences underlying this topic and mental health in general (Bruine de Bruin 2020). It's main objective was to understand age differences in risk perceptions, anxiety and depression using a nationwide representative survey to 6,666 US adults from March 2020. It started by presenting two behavioural theories in this regard: Socio-emotional Selectivity Theory (Carstensen 2006) and the Strength and

[2] "0 = strongly disagree/disagree/neutral; 1 = agree/strongly agree"

Vulnerability Model (Charles 2010). This survey associated older age with the perception of a greater infection-fatality risk while linking such seniority with "seeing lower risks of getting COVID-19 and of experiencing negative economic effects" and with lower levels of depression and anxiety. The paper suggests that this lower reporting of negative emotions and of responses to daily stressors might be explained through some kind of emotional regulation mechanism used to focus on "the bright side" or to choose stress-reducing activities in these situations.

In Carstensen (2006), the author uses Socio-emotional Selectivity Theory (SST) to attempt to explain age differences like the ones previously exposed. The main takeaway of the paper is that, as an individual gets older, the concept of remaining time in their lives becomes a better predictor than chronological age for several mental, motivational and emotional variables. It suggests that this variance is mainly put forwards by differences in experiences and opportunities that any given person has lived through at a specific point in time. This is due to a preference trade-off between the acquisition of knowledge and the regulation of one's emotional state. More precisely, the paper posits, on the one hand, that at a younger age, when one perceives time as virtually endless, their main goals are related to the preparation of what comes ahead and to acquiring knowledge on the world around you in order to conquer long-term achievements. On the other hand, at an older age, when life is believed to be ending relatively soon, one tends to prioritize short-run goals and anything that can lead them to maximize their psychological well-being in the short time they have left. The paper finishes by looking at yet another age difference in perception, now related to comparing positive and negative stimuli, in order to understand if a particular age group presents a specific attraction to any of the two types. The paper suggests that younger generations tend to find negative information more compelling and "attention-grabbing", while the elderly seem to prefer to remain optimistic and to focus on positive information. This shift in priorities might have a role to play in the determination of risk perception and their impact on human behaviour.

The Strength and Vulnerability Integration (SAVI) Model (Charles 2010) takes on this theory and attempts to determine the mechanisms that influence emotional regulation and well-being at different life stages. It argues that age is related to an increase in emotional strength to deal with stressors both related to both present and past events. It uses SST to suggest that this emotional resilience might arise from a change in perspective provided by the fact that older people worry more about the time they have left to live and not so much about the amount of years they have lived so far. In other words, it uses SST to indicate that the motivation behind the adoption of coping mechanism may come from the joint force of a young-adult focused on the short-run and on prioritizing the collection of new experiences and information, regardless of the possible negative emotional impacts of such situations, and an older adulthood spent worrying about the time left on this world and the right way to maximize their own psychological well-being. Thus, when older individuals can manage to use these experiences to adapt themselves and employ behavioural strategies to minimize an emotionally negative situation, they can in fact regulate their emotional state and improve their affective well-being.

More papers followed suit with trying to shed a light on the new Coronavirus. Dryhurst et al. (2020) deemed clear that the spread of the virus was influenced by the public's risk perception on the topic and their willingness to take preventive measures against it. The paper identified some factors that could influence one's expectation of the probability of being infected and negatively impacted by the SARS-CoV-2 virus, namely first-hand experience with it, individualistic or prosocial values, the source of information on the topic, and the effectiveness of the implemented measures to fight against the pandemic. These were aggregated into four main groups: cognitive tradition (comprising one's knowledge on risk); experiential tradition (related to personal experiences); the socio-cultural paradigm (how the society in which an individual is inserted views risk, their values and cultural ideals); and any relevant individual differences (a translation of a person's demographic profile in regards to gender, education, occupation, ideology, among others). The authors also point out the importance of accurately perceiving both the personal and communitarian risk associated with COVID-19 for the success of any policy or strategy to minimize the devastating effects of this crisis. They find this point so important that they use the Protection-Motivation Theory (Floyd et al. 2000) as a framework to understand the threat analysis process that determines one's cooperation with any kind of restrictive or precautionary behaviour. In this way, one can better advise policy-makers on what to have in mind in order to maximize the effectiveness of any measure to fight against the pandemic and get people to cooperate and work together on a solution.

The Protection Motivation Theory (PMT), looked at under the scope of a meta-analysis by Floyd et al. (2000), intends to set up a framework to uncover what motivates a person to accommodate policies that implicate some kind of restriction on their lives rather than just taking the risk of being negatively affected by what those policies are attempting to contain. From a set of four theories, which all share the idea that protection motivation derives from a cost-benefit analysis of a given threat and one's willingness to prevent the possible negative effects associated with that menace, the authors decided to focus on PMT. This choice is justified by the importance given by the model to the individual notion of the effectiveness of the preventive behaviour to be undertaken and to its format that encompasses two processes to simulate human threat evaluation. This theory posits that, when faced with a possible threat and the chance to perform a certain preventive action, people collect information from both their environment and their own experiences in order to perform a cost-benefit analysis of the threat and the specific coping mechanism, coming up with their own protection motivation that dictates whether they take an adaptive or maladaptive coping stance. This cost-benefit evaluation weights, for the threat appraisal, the rewards of not adopting the coping mechanism against the severity of their vulnerability regarding the threat in question, and, for the coping appraisal, the responses' efficacy against the cost of putting that mechanism in place. This paper establishes a possible direction regarding the chain of events that determines someone's response to COVID-19, or lack thereof.

The number of papers referencing the current pandemic only grew larger as researchers tried to do their best to use their expertise to help Humanity surpass this crisis in any way they could. Breznau (2020) attempts to understand the role of welfare states on a global emergency. It uses comparative data on 70 countries, from April 2020, to attempt to establish a relationship between the strength of a welfare state and the risk perceptions of those under its domain. In order to measure risk perception, this paper uses five questions regarding the respondents' concerns about the consequences of the SARS-CoV-2 on themselves, on their families, friends, countries, and on other countries, giving a large span of perspectives on the threat. Furthermore, this paper takes into account the number of days since the inflexion of the curve depicting the evolution of new cases and the number of new cases in the week prior to the surveys in order to capture the long and short-term repercussions of those notions, respectively, on the effectiveness of State measures to control the pandemic. It also looks at the effect of media on risk perceptions as it understands that this factor influenced, at the time, the public's knowledge on the matter and on the effectiveness of the measures put in place. In it, the author claims to demonstrate that the stronger a given welfare state is, the more it is associated with lower risk perceptions due to its role in providing "social insurance and protection policies" to the public.

These materials serve to help creation a complete possible framework of how the information and policies provided by the State, and the Portuguese government in particular, can result in certain behaviours, coping or otherwise, on the behalf of the general population. The first step was accomplished by reviewing papers like McFadden et al. (2020) that provided a quick snapshot of how the crisis was perceived at a time with an extremely low information available, for example, about what could be an expectable distribution of the risk perceptions or the main information providers. After this initial notion, some papers like Carstensen (2006), Charles

(2010) and Floyd et al. (2000) proved useful in providing a possible skeleton for the whole procedure, a way to connect stimuli, the formation of a risk perception, and the usage of that information to make a decision. The first presents a framework to justify age differences in risk perceptions, while the second tries to build upon this idea so as to create a more comprehensive model and the third argues in favour of a framework for a cost-benefit analysis of a given threat and its coping mechanisms. Lastly, Breznau (2020), Dryhurst et al. (2020) and Bruine de Bruin (2020) helped revealing the fundamental and most salient factors that determined an individual's perception of risk concerning this global epidemic. Variables suggested by these papers that are used below are the number of new cases in the week prior to the surveys, age in categorical intervals and pre-existing conditions as part of one's experiential tradition.

3. Data

The current study was performed mainly using a survey designed at the Nova Health Economics Knowledge Centre and field works carried out by the company GFK in Portugal between the 23rd of May and the 30th of June 2020 in relation to the Chair BPI | Fundação "La Caixa" in Health Economics. This survey, hereafter referred to as GFK2020, was handed out to individuals aged 15 or older residing in Continental Portugal, constituting a sample of 1,271 interviews throughout the whole country. The interviewees were chosen considering a system of quotas that crossed Region (7 groups), Habitat (5 groups), Gender (2 groups), Age (6 groups), Schooling (2 groups) applied to men and Occupation (2 groups) applied to women. From an initial matrix crossing Region and Habitat, a significant number of sampling points was randomly selected, and interviews were performed on those points considering those quotas. The surveys were conducted through an in-person interview using a Computer Assisted Personal Interviewing (CAPI) system in the interviewees' homes to guarantee full privacy.

The survey consisted of a series of questions regarding individual habits,[3] medical conditions, and some demographic parameters, among many others. The central question asked how the interviewee felt regarding the risk associated with the spreading of COVID-19, applied to four different levels of proximity to the individual. These risk perception measures should be carefully analysed as they represent reported data and not an objective measure of risk. Thus, any conclusion concerns only the individual expectation of risk and not necessarily the actual risk a person is subject to because of the Coronavirus pandemic. Additionally, binary forms of the dependent variables were generated by setting the value 1 for risk evaluations equal of higher than 3, thus transmitting a notion of neutral or high risk, and value 0 for perceptions below 3, thus meaning low risk. Other transformations were performed in variables associated with age and number of household members to make them easier to use in the models.

Variable	**Question**
Average Daily % Change in the Number of COVID Cases	The arithmetic average of the percentage change in the cumulative number of cases per region in the week prior to the survey.[4]
Pre-Existing Conditions	"Do you have any chronic condition that requires specific medication?"
Asthma	"Could you tell me which disease(s) is/are that/those?"
Lung Disease	
Heart Attack	
Heart Disease	
Hypertension	
Stroke	
Arthritis	
Back Pain	
Neck Pain	
Diabetes	
Allergies	
Hepatic Cirrhosis	
Bladder Control	
Kidney Disease	
Depression	
Cancer	
High Cholesterol	

[3] For information on the questions behind each variable and their descriptive statistics, see Figure 1 in Annex I, Table 14 through 58 in Annex II, and Figure 11 also in Annex II.
[4] See Equation 1 in Annex III

General Risk Individual Risk Family Risk Community Risk	“We would like to understand what you think and feel regarding the risk associated with the spreading of COVID-19. Please, use a scale to evaluate your probability of being infected with the new Coronavirus and its possible consequences. To do so, use a scale from 1 to 5 in which ‘1’ means ‘No risk’ and ‘5’ means ‘Very high risk’. You can use any intermediate value to provide us with your evaluation.”: • “My risk of getting infected with the Coronavirus” (General Risk) • “The risk for my health due to COVID-19” (Individual Risk – High Proximity) • “The risk for the health of my family’s members due to COVID-19” (Family Risk – Medium Proximity) • “The risk for the health of my community’s members due to COVID-19” (Community Risk – Low Proximity)
Private Health Insurance	“Do you have any private health insurance?”
Smoke	“Which of the following statements best describes your situation regarding tobacco consumption? Would you say you…”
Drink	“Which of the following statements best describes your situation regarding the consumption of alcoholic beverages? Would you say that you consume alcoholic beverages…”
General Health State	“In general, how would you consider the state of your health? Would you that it is…”
Gender	“Register Gender”
Age	“Please, tell me, how old are you?”
Household Members	“How many people live in your house, including yourself, both adults and children?”
Education	“And what is the highest level of instruction you have completed?”
Age by Interval	Categorical version of the variable Age comprised of 7 brackets.
Household Members by Interval	Categorical version of the variable Household Members comprised of 3 brackets.

Table 1 - Description of the main variables included in the models

As suggested in Breznau (2020), a variable containing the average rate of change of the number of COVID cases during the week prior to the interviews in each region was integrated in the paper so as to capture a short-term effect of COVID cases on all measures of risk perception. The work-project also uses a “yes” or “no” question on whether the respondent had a specific chronic condition and questions regarding smoking (ranging from 0 to 3, with the former meaning that the individual smoked daily and the latter that they have never smoked), alcohol consumption (ranging from 0 to 8, with the former meaning that they consume alcohol on a daily or almost daily basis and the latter meaning that they have never consumed any or have only tasted it occasionally) and the respondent’s views on their general health state (ranging from 0 to 4, with 0 meaning “Very good” and 4 meaning “Very bad”). Furthermore, a binary variable representing if a given individual reported having a private health insurance scheme or not was also included, as well as some demographic parameters such as gender (as a binary variable establishing the baseline as a male individual), age (transformed into a categorical variable ranging between 0, the younger, and 6, the older), number of household members (transformed into a categorical variable ranging between 0, small households, and 2, large households), education level (a categorical variable representing 8 possible levels of instruction, the lowest being the level 7 and the highest the level 0), occupation (a categorical variable ranging from 0 to 5, with each level representing a given occupation), and the region (ranging from level 0 – Northern Seaside – to 6 – Algarve) in which they reside. All these variables are used as controls in an attempt to understand what determines a person’s perception of risk regarding COVID-19 based on health determinants that could influence their vulnerability to the disease, and on their personal information.

The GFK2020 database also presented respondents with another question related to their precautionary behaviours associated with COVID-19. This generated binary variables concerning whether the interviewee, at the time of the survey and as an attempt to mitigate the spread of the virus, had started washing their hands more frequently, wearing a mask in public, not shaking hands, not going outside at all, among many other responses. However, these factors were not included in the project to prevent problems of reverse causality in parameter estimation. In other words, a person can reduce their social contacts or adopt preventive behaviours due to their high perception of the risk, or, conversely, one can have their notion of risk be impacted by their precautionary behaviours, i.e. if someone starts covering their face at all times, then they might feel safer and thus perceiving a lower risk associated with the spread and negative consequences of the new Coronavirus.

Nonetheless, risk perception is an extremely complex and personal concept that is not easily explained by the usual explanatory variables. Although a person's gender, occupation and income level might influence the way they perceive the world and its threats,[5] that individual also has a set of experiences, a background, that plays an important role in determining what they value the most, what their willing to risk and, thus, their particular notion of risk regarding COVID-19. Not only that, but so does the threat itself, how it is impacting the world right now, how it has impacted one's environment so far, what kind of restriction it has imposed on their life, among many others. Thus, as in previous literature, there could be value added to this research in including the number of days since the inflection of the epidemic curve (Breznau, 2000) at the time of the survey, as a way of providing a long-term view on the crisis since the further one is past this point, the more probable it could be for a flattening of the curve to be seen as a success and the lower could one's risk perception be. It could also be the case that during this time people got used to life in a pandemic crisis, thus changing their reference point regarding the normal life standard, which could translate into a lower risk perception. Finally, the same can be said about a variable representing the movement restrictions in place at that time. All these variables, and others related to cultural differences, cognitive capabilities and other parameters, could have been important additions so as to use variables connected to COVID-19 and non-demographic information as explanatory variables instead of only trying to explain risk perception on this topic through health-related and demographical parameters.

[5] Initially, the intention of the paper was to look at the effect of income on the chosen dependent variables. However, after testing for the existence of endogeneity and some kind of selectivity problem in terms of those that chose to report their income level, this variable had to be withdrawn from the study.

4. Methodology

The first step is to analyse the GFK dataset, apply the necessary transformations to the available variables, and determine the main variables of interest. In this study, the dependent variables are the risk perceptions at individual, familiar and communitarian levels.[6]

The main information under the scope was risk perception and how likely a person is of perceiving a higher level of risk with regards to COVID-19, which is transmitted by the binary form of the variables related to the notion of danger to one's health (*Individual Risk*), to one's family's health (*Family Risk*), and to one's community's health (*Community Risk*). It is worth pointing out that the defined variables of interest are interconnected with each other, i.e. a person's family's externalities add on to their individuals expectations to form their family-level risk notion concerning COVID-19 and social externalities add on to these expectations to form their community-level risk notion. This means that, most likely, any possible information that is not captured by the existing covariates will also be extremely correlated. This is an important point to make as this project intends to understand the determinants of these three notions and how the former impact the latter, meaning that it is crucial to account for this notion of interdependence so as to help the models better predict the true effects under study.

Thus, the models developed in this study use two basic ideas to justify their formulation: The utilization of probability models to focus on the chance of a given individual perceiving a relatively higher risk related to the Coronavirus, thus the binary forms; The interdependence among the dependent variables, which makes it appropriate to attempt to regress the models of the three selected notions of risk simultaneously so as to guarantee that their error terms reflect, for example, the similar impact of information shocks on them. Thus, three different general models are used in order to achieve these goals and infer on the determinants of COVID-19 risk

[6] Descriptive statistics in Tables 41 through 43 in Annex II

perception: Linear Probability Model (LPM);[7] Maximum Likelihood Estimation (MLE) Model under a Probit functional form;[8] Seemingly Unrelated Regressions (SUR) Model, which extends on the simple LPM.[9]

The first assumes a linear model under the usual assumptions necessary for the conduction of valid inference, which allows for an interpretation of the estimated coefficients as constant, average, and *ceteris paribus* effects of the chosen covariates on a specific measure of risk perception. However, due to its linearity, it is possible for probabilities over 1 and below 0 to exist, which is not compatible with standard probabilistic notions. Nonetheless, its simplicity might still be useful if the Probability Distribution in question does not present such extreme values. Thus, this type of approach requires that heteroskedasticity be addressed to satisfy these assumptions, which is done by computing the Breusch-Pagan and White tests for heteroskedasticity,[10] leading to the maintenance of of the null hypothesis of homoskedasticity in all risk notions except for *Community Risk*, and the use of robust standard errors in the latter equation. Furthermore, in order to assess the reliability of assuming that the error terms present an asymptotically Normal Distribution, histograms for the residuals of the three robust regressions were plotted against the Normal Distribution,[11] presenting a relatively consistent distribution that allows the author to retain that assumption, with due caution.

The second approach involved the use of the MLE method, which computes the coefficients that maximize the probability of observing a given dependent variable associated to the selected covariates. In this case, the approach uses a Probit model, meaning that it is assumed that the Cumulative Distribution of the dependent variables conditional on other covariates follows a Standard Normal Cumulative Distribution. This method’s upside is that, by taking a non-linear

[7] Equations 2, 3 and 4 in Annex III
[8] Equations 2, 3 and 4 in Annex III
[9] System composed of Equations 2, 3 and 4 in Annex III
[10] Tables 7 through 12 in Annex I
[11] Figures 2, 3 and 4 in Annex I

approach and by assuming a certain probability distribution, it forces those probability values to be contained between 0 and 1, thus becoming ideal for situations in which they are either too high or too low and need to be restricted. However, the MLE is more complex than the LPM, meaning that since the model does not take a linear approach, the effect of a covariate on a dependent variable is dependent on the specific value of the covariate. In other words, to be able to compare estimates in different models one has to use Marginal Effects (ME) instead of simply looking directly at the coefficients in question. Thus, by computing the ME at the mean value of all variables one can obtain a somewhat on average, *ceteris paribus*, evaluation of their effects that is comparable to those estimated through LPM and SUR.

The SUR method is based on the idea that the three dependent variables are subject to the same random shocks. This model works by estimating all equations as a system to account for their contemporaneous cross-equation error correlation. This process will not output different regression coefficients, or *R-squares*, but will influence standard errors, i.e. it will compute different estimations that account for the correlation between the three equations and that might determine a different set of statistically significant variables than a regular LPM.

Apart from understanding whether this treatment differs significantly from a regular LPM, it will also be important to look at differences in the probabilities of having a higher risk perception concerning COVID-19 themselves.

During the formulation of this project's methodology, a multiple regression study on geographical mortality (Cook and Pocock 1983) that suggests an approach to account for a correlated error structure of mortality rates in geographically close areas that might be similar for motives not captured by the known explanatory variables, meaning that their errors could be dependent of each other, which could lead to a serious overestimation of the established relationships was analysed in order to understand what possible procedures could be used to account for such an effect. The authors try to model this spatial correlation in an example using data from the British Regional Heart Study to support their method instead of simply aggregating many small units into fewer larger ones that incorporate the regional similarities, preventing them from posing a threat to the validity of this type of studies. Nonetheless, after much consideration, the present methodological approach ended up being chosen.

5. Results

The probability of a respondent presenting a relatively high risk perception regarding COVID-19 is greater at an individual level (with a probability of 43.8% and 44.8% according to LPM and SUR or MLE, respectively) than at a family level (42.5% and 41.9%), but that same probability at a communitarian stage is higher than any of the other measures (with a probability of 51.3% independently of the regressing procedure). This is worrisome to the extent that only slightly more than one in every two individuals perceive a high *Community Risk* regarding the new Coronavirus, and even less do so concerning the other levels of risk. This result may be a sign that contention measures could be harder to implement due to an underestimation of the risk intrinsic to the virus at all these different levels of proximity.

Concerning the chosen independent variables,[12] firstly, the average percentage change in the number of cases during the week immediately prior to the survey seems to be associated with an increase in the probability of perceiving a high risk related to COVID-19 (except for *Community Risk*), which supports previous literature in their argument for the importance of the number of cases in supporting the heightening of risk perceptions in this pandemic crisis. This duality reveals an important policy implication since whether the adoption of precautionary measures is primarily determined by an individual or communitarian notion of risk would impact the direction of the impact of this information, i.e. if there is an altruistic

[12] Complete analysis in Annex IV

preoccupation with the well-being of one's peers, then it would make them more likely to perceive a higher risk-level, while if this individual is more egocentric, then it would make them more prone to perceive a lower risk-level.

Furthermore, having one or several respiratory diseases could be expected to increase the chance that someone has an elevated risk perception for their individual and family health due to the frailties created in the organs most affected by the virus. However, at a community level, conditions like Lung Disease, Heart Attack and Heart Disease are found to decrease this probability. Nonetheless, Diabetes, Asthma (only for *Individual Risk*) and Allergies present statistically significant relationships with the dependent variables, which is particularly important for the last two diseases since they affect the respiratory system, the organs most affected by the virus.

This concern with the major organs affected by the virus could also lead to the expectation that as one smokes more frequently, harming one such organ and thus increasing their vulnerability to COVID-19, they could be more prone to internalize that possibility and be associated with an increase in this probability. In fact, this is the case with all three risk notions, which is relevant in indicating that people seem to be aware of the negative impact of smoking and to be internalizing that knowledge into their risk notions at all proximity levels.

There are also other destructive behaviours that could affect one's notion of risk regarding the new Coronavirus, such as alcoholic consumption, which affects a person both physically and psychologically. In this regard, data seems to indicate that a highly frequent consumption is associated with an increase in the probability of perceiving a high level of *Individual* and *Family Risk*, but not of *Community Risk*. This duality might indicate some kind of difficulty in processing cognitively the impact of individual and family habits on communitarian health.

These notions help shape one's idea of their personal health, which could be expected to connect with their risk perception in such a way that a worse health state could be associated with an increase the probability of perceiving a high risk. With the available data, there is some support for this statement at an individual and family levels, although not for *Community risk*, which depicts a much more stable evolution of the impact in these risk perceptions with the worsening of one's reported general health state.

Nonetheless, in a pandemic where close contacts must be minimized, it is not only internal factors that count but also the size of one's bubble of regular close contacts. In this sense, a higher number of household members could be expected to increase the chance of them conserving a high risk perception exactly because of the need to keep social distancing and the fact that that separation becomes easier with a small and closed set of family members living in the same quarters. However, data seems to support this expectation at all levels except for *Family Risk*, in which a larger household is associated with a decrease in this probability at all levels of proximity, which contradicts the previously mentioned expectation, but could be a sign of a notion that large families see themselves as constrict circles of close contacts, capable of isolating itself from the outside and attain some kind of group immunity.

Some literature reviewed for this paper pointed towards the possible adoption of coping mechanisms by individuals of an older age in order to mitigate the negative effects of daily stressors and maximize their psychological well-being for the remainder of their life. The present data did not support this notion, presenting a clear tendency of older age intervals to be associated with a higher increment in the probability that a given person perceives a high risk regarding the virus, even if in slightly less pronounced manner for *Community Risk*. Thus, the present results seem to confirm the pre-existing notion that the fact that older people are more vulnerable to the disease makes the more likely to be more preoccupied with it.

Finally, the most interesting results to point out concern the indication of the possible existence of a self-selection situation at all levels in the variables of *Private Insurance* and *Education*. In the first case, those with private health insurance are associated with an increase in the chance of perceiving a high risk, which might mean that those with higher risk perceptions could be the ones that search the most for private health insurance in the first place. In the last case, more educated individuals are associated with a higher increase in the probability of perceiving a high risk related to COVID-19 (except at a communitarian level), which might be a signal that they are more attentive to these kinds of matters, or that they try harder to search and become more knowledgeable on them, or that they use their skills to search for higher-quality information that directs their notions, thus putting themselves on a situation that can make them more prone to perceive a higher *Individual* and *Family Risk*. In these two cases, there is space for a factor of adverse selection or even self-selection to play a role on risk perceptions and in pushing these groups towards an accurate notion of risk. In other words, these results represent patterns that, together with the remaining results, can help design communication policies for the duration of the pandemic crisis and future threats.

This empirical evidence has real world effects on the adoption of precautionary measures and the minimization of the negative consequences deriving from the current pandemic crisis. It is fundamental that societies perceive the correct level of risk in order to perform an accurate cost-benefit analysis and employ the necessary effort to fight against the virus. Thus, by creating campaigns and other methods to spread information targeting those individuals most likely to perceive a low level of risk, States can nudge them towards taking action and preventing an underestimation of the nefarious consequences that might fall upon them.

	Individual Risk			**Family Risk**			**Community Risk**		
	LPM	**MLE's ME**	**SUR**	**LPM**	**MLE's ME**	**SUR**	**LPM**	**MLE's ME**	**SUR**
Average Daily % Change in the Number of COVID Cases	1.968	2.158	1.968	0.603	0.120	0.603	-3.082	-3.331	-3.082
	(0.584)	(0.632)	(0.574)	(0.868)	(0.978)	(0.865)	(0.446)	(0.438)	(0.422)

Asthma	0.145	0.239*	0.145	0.152	0.264*	0.152	0.106	0.128	0.106
	(0.125)	(0.100)	(0.115)	(0.114)	(0.066)	(0.104)	(0.287)	(0.301)	(0.294)
Lung Disease	0.039	0.067	0.039	0.115	0.182	0.115	-0.119	-0.182	-0.119
	(0.708)	(0.661)	(0.701)	(0.274)	(0.241)	(0.261)	(0.291)	(0.179)	(0.283)
Heart Attack	0.147	0.187	0.147	0.049	0.030	0.049	-0.047	-0.112	-0.047
	(0.309)	(0.455)	(0.296)	(0.736)	(0.882)	(0.730)	(0.741)	(0.563)	(0.761)
Heart Disease	0.031	0.033	0.031	0.107	0.167	0.107	0.007	-0.001	0.007
	(0.775)	(0.835)	(0.769)	(0.337)	(0.301)	(0.325)	(0.952)	(0.993)	(0.953)
Hypertension	0.060	0.100	0.060	0.033	0.049	0.033	0.014	0.006	0.014
	(0.245)	(0.190)	(0.233)	(0.524)	(0.472)	(0.513)	(0.798)	(0.933)	(0.805)
Stroke	0.010	-0.099	0.010	-0.026	-0.041	-0.026	0.071	0.200	0.071
	(0.946)	(0.709)	(0.944)	(0.865)	(0.871)	(0.862)	(0.484)	(0.468)	(0.659)
Arthritis	0.023	0.120	0.023	-0.118	-0.163	-0.118	-0.068	-0.111	-0.068
	(0.774)	(0.400)	(0.768)	(0.153)	(0.141)	(0.142)	(0.405)	(0.312)	(0.436)
Back Pain	-0.008	0.053	-0.008	0.090	0.173	0.090	0.056	0.140	0.056
	(0.930)	(0.723)	(0.928)	(0.334)	(0.204)	(0.321)	(0.473)	(0.304)	(0.564)
Neck Pain	0.030	0.042	0.030	-0.142	-0.268*	-0.142	-0.073	-0.133	-0.073
	(0.804)	(0.823)	(0.799)	(0.246)	(0.092)	(0.233)	(0.542)	(0.405)	(0.572)
Diabetes	0.174***	0.352***	0.174***	0.131**	0.211***	0.131**	0.171***	0.242***	0.171***
	(0.001)	(0.000)	(0.001)	(0.016)	(0.005)	(0.013)	(0.001)	(0.001)	(0.003)
Allergies	0.308**	0.394**	0.308**	0.477***	0.647***	0.477***	0.216*	0.282*	0.216
	(0.013)	(0.014)	(0.011)	(0.000)	(0.000)	(0.000)	(0.077)	(0.089)	(0.103)
Hepatic Cirrhosis	-0.099	-0.112	-0.099	-0.052	-0.080	-0.052	-0.084	-0.141	-0.084
	(0.574)	(0.672)	(0.564)	(0.770)	(0.752)	(0.764)	(0.628)	(0.535)	(0.655)
Bladder Control	-0.048	0.000	-0.048	0.106	0.000	0.106	0.301**	0.000	0.301
	(0.822)	(.)	(0.817)	(0.624)	(.)	(0.615)	(0.016)	(.)	(0.188)
Kidney Disease	0.075	0.058	0.075	0.098	0.052	0.098	0.201*	0.352	0.201
	(0.584)	(0.775)	(0.575)	(0.482)	(0.771)	(0.471)	(0.055)	(0.152)	(0.170)
Depression	-0.004	0.001	-0.004	-0.021	-0.019	-0.021	-0.039	-0.034	-0.039
	(0.958)	(0.995)	(0.957)	(0.798)	(0.851)	(0.793)	(0.654)	(0.736)	(0.649)
Cancer	0.106	0.087	0.106	0.030	0.003	0.030	-0.048	-0.094	-0.048
	(0.422)	(0.662)	(0.410)	(0.822)	(0.985)	(0.817)	(0.707)	(0.584)	(0.734)
High Cholesterol	-0.034	-0.044	-0.034	0.015	0.039	0.015	0.029	0.027	0.029
	(0.896)	(0.910)	(0.893)	(0.955)	(0.912)	(0.953)	(0.873)	(0.938)	(0.917)
Private Health Insurance	0.007	0.015	0.007	0.005	0.010	0.005	0.031	0.034	0.031
	(0.860)	(0.764)	(0.857)	(0.901)	(0.844)	(0.898)	(0.466)	(0.470)	(0.453)
Smoke Occasionally	0.165**	0.210**	0.165**	0.009	0.035	0.009	0.017	0.028	0.017
	(0.045)	(0.032)	(0.040)	(0.919)	(0.733)	(0.916)	(0.856)	(0.773)	(0.845)
Have smoked, but currently does not smoke	0.031	0.039	0.031	0.019	0.023	0.019	-0.059	-0.066	-0.059
	(0.538)	(0.546)	(0.527)	(0.702)	(0.714)	(0.694)	(0.281)	(0.274)	(0.263)
Never Smoked	0.003	-0.005	0.003	0.003	-0.004	0.003	-0.043	-0.050	-0.043
	(0.940)	(0.912)	(0.939)	(0.942)	(0.934)	(0.940)	(0.282)	(0.233)	(0.253)
Drink between 5 to 6 days a week	0.004	0.015	0.004	0.059	0.074	0.059	-0.070	-0.096	-0.070
	(0.949)	(0.870)	(0.948)	(0.403)	(0.408)	(0.391)	(0.332)	(0.254)	(0.344)
Drink between 3 to 4 days a week	-0.020	-0.018	-0.020	-0.004	-0.004	-0.004	0.095*	0.111*	0.095*
	(0.700)	(0.792)	(0.692)	(0.946)	(0.947)	(0.944)	(0.093)	(0.083)	(0.091)
Drink between 2 to 1 days a week	-0.121**	-0.157**	-0.121**	-0.050	-0.064	-0.050	-0.038	-0.042	-0.038
	(0.017)	(0.013)	(0.015)	(0.330)	(0.308)	(0.317)	(0.503)	(0.495)	(0.485)
Drink between 2 to 3 times a month	-0.039	-0.033	-0.039	0.048	0.072	0.048	0.091	0.106*	0.091*
	(0.435)	(0.600)	(0.422)	(0.346)	(0.252)	(0.333)	(0.111)	(0.077)	(0.090)
Drink once a month	-0.039	-0.046	-0.039	-0.048	-0.058	-0.048	0.088	0.100	0.088
	(0.613)	(0.646)	(0.604)	(0.538)	(0.561)	(0.528)	(0.301)	(0.277)	(0.280)
Drink less than once a month	-0.042	-0.052	-0.042	-0.060	-0.068	-0.060	0.045	0.053	0.045
	(0.404)	(0.428)	(0.392)	(0.244)	(0.284)	(0.232)	(0.445)	(0.388)	(0.413)
Has not consumed alcohol in the last 12 months, due to having stopped drinking	-0.010	0.000	-0.010	-0.019	-0.027	-0.019	0.081	0.093	0.081
	(0.883)	(0.997)	(0.880)	(0.775)	(0.752)	(0.769)	(0.262)	(0.250)	(0.251)
Never consumer alcohol, or just	0.015	0.025	0.015	-0.014	-0.013	-0.014	0.044	0.051	0.044

occasionally tasting it									
	(0.739)	(0.680)	(0.732)	(0.759)	(0.821)	(0.753)	(0.401)	(0.365)	(0.376)
Good General Health State	0.025	0.022	0.025	-0.047	-0.069	-0.047	-0.030	-0.039	-0.030
	(0.506)	(0.623)	(0.495)	(0.214)	(0.130)	(0.202)	(0.475)	(0.377)	(0.455)
Reasonable General Health State	0.239***	0.244***	0.239***	0.155***	0.148**	0.155***	-0.006	-0.018	-0.006
	(0.000)	(0.000)	(0.000)	(0.001)	(0.010)	(0.001)	(0.912)	(0.747)	(0.906)
Bad General Health State	0.371***	0.560***	0.371***	0.298***	0.415***	0.298***	0.208***	0.311***	0.208**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.001)	(0.019)
Very Bad General Health State	0.440	0.000	0.440	0.287	0.000	0.287	-0.113	-0.122	-0.113
	(0.117)	(.)	(0.107)	(0.312)	(.)	(0.299)	(0.627)	(0.749)	(0.706)
Gender (Male as Baseline)	0.033	0.046	0.033	0.053*	0.074**	0.053*	0.016	0.022	0.016
	(0.242)	(0.215)	(0.230)	(0.063)	(0.040)	(0.057)	(0.606)	(0.519)	(0.591)
Ages between 25 and 34 years old	0.116**	0.142**	0.116**	0.119**	0.147**	0.119**	0.124*	0.131*	0.124**
	(0.045)	(0.045)	(0.040)	(0.043)	(0.031)	(0.037)	(0.054)	(0.054)	(0.044)
Ages between 35 and 44 years old	0.125**	0.158**	0.125**	0.155***	0.195***	0.155***	0.159**	0.172**	0.159***
	(0.030)	(0.023)	(0.026)	(0.008)	(0.004)	(0.006)	(0.015)	(0.010)	(0.009)
Ages between 45 and 54 years old	0.105*	0.127*	0.105*	0.102*	0.136*	0.102*	0.128*	0.145**	0.128*
	(0.086)	(0.085)	(0.078)	(0.100)	(0.056)	(0.091)	(0.062)	(0.044)	(0.050)
Ages between 55 and 64 years old	0.194***	0.219***	0.194***	0.207***	0.237***	0.207***	0.169**	0.188**	0.169**
	(0.004)	(0.007)	(0.003)	(0.002)	(0.003)	(0.002)	(0.023)	(0.018)	(0.019)
Ages between 65 and 74 years old	0.284***	0.376***	0.284***	0.258***	0.318***	0.258***	0.265***	0.335***	0.265***
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.003)	(0.005)	(0.001)	(0.005)
More than 75 years old	0.274***	0.338***	0.274***	0.254***	0.290**	0.254***	0.178*	0.182	0.178*
	(0.004)	(0.006)	(0.003)	(0.008)	(0.016)	(0.006)	(0.073)	(0.123)	(0.078)
Between 3 and 4 household members	0.006	0.014	0.006	-0.015	-0.017	-0.015	-0.015	-0.014	-0.015
	(0.826)	(0.718)	(0.822)	(0.616)	(0.646)	(0.607)	(0.649)	(0.703)	(0.631)
5 or more household members	-0.029	-0.034	-0.029	-0.021	-0.026	-0.021	0.012	0.011	0.012
	(0.570)	(0.610)	(0.560)	(0.681)	(0.687)	(0.673)	(0.830)	(0.855)	(0.826)
Middle Course/ Polytechnic Degree	0.178	0.214	0.178	0.214*	0.249*	0.214*	0.030	0.034	0.030
	(0.140)	(0.133)	(0.130)	(0.081)	(0.084)	(0.073)	(0.820)	(0.811)	(0.814)
Participation in Higher Education/ Middle	0.190*	0.239**	0.190**	0.121	0.153	0.121	0.113	0.128	0.113
	(0.054)	(0.044)	(0.048)	(0.225)	(0.216)	(0.213)	(0.270)	(0.287)	(0.282)
12th Grade (7th Year of High School/ 11th Grade)	0.010	0.010	0.010	0.036	0.042	0.036	0.045	0.053	0.045
	(0.843)	(0.871)	(0.839)	(0.459)	(0.461)	(0.447)	(0.402)	(0.362)	(0.389)
9th Grade (5th Year of High School)	0.065	0.088	0.065	0.113**	0.148**	0.113**	0.118**	0.140**	0.118**
	(0.207)	(0.180)	(0.196)	(0.031)	(0.017)	(0.026)	(0.038)	(0.025)	(0.034)
6th Grade (2nd Year of High School)	0.057	0.063	0.057	0.098*	0.111	0.098*	0.085	0.096	0.085
	(0.329)	(0.395)	(0.316)	(0.097)	(0.111)	(0.089)	(0.177)	(0.172)	(0.170)
Complete Primary Education	0.078	0.104	0.078	0.140**	0.181**	0.140**	0.085	0.094	0.085
	(0.187)	(0.171)	(0.176)	(0.020)	(0.012)	(0.017)	(0.187)	(0.195)	(0.178)
Incomplete Primary Education/ Analphabet	0.075	0.073	0.075	0.058	0.032	0.058	0.225***	0.310***	0.225**
	(0.388)	(0.574)	(0.375)	(0.508)	(0.765)	(0.497)	(0.005)	(0.004)	(0.015)
Constant	0.158*	0.000	0.158*	0.227**	0.000	0.227**	0.276**	0.000	0.276***
	(0.096)	(.)	(0.087)	(0.019)	(.)	(0.016)	(0.011)	(.)	(0.007)
Observations	1,230	1,222	1,230	1,230	1,222	1,230	1,230	1,225	1,230
R^2	0.285		0.285	0.261		0.261	0.153		0.153
Adjusted R^2	0.248	0.155	0.248	0.222	0.127	0.222	0.108	0.046	0.108
Pseudo R^2		0.239			0.211			0.128	

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2 - Main coefficients estimated through all models at all three levels of proximity

6. Conclusion

This paper attempts to shed light on the determinants associated with a higher chance of perceiving a high risk related do the new Coronavirus, so as to understand how that might impact their cooperation with all the measures put in place to contain the spread of the virus.

Departing from the result indicating that, at the mean values of all its covariates, the predicted probability for the binary form of *Community Risk* is higher than that of *Individual Risk*, which in turn is higher than that of *Family Risk*, one starts to have an idea of how the perception of risk has a higher probability of being high at an communitarian level, i.e. the possibility of making other suffer from the new Coronavirus speaks louder than that of themselves suffering.

The truth is that most estimators were not statistically significant in all three models. However, there could still exist some relevant associations to keep in mind. First of all, the fact that having private health insurance increases the chance of someone having a high risk perception might be a sign that those with a higher risk perceptions could be the ones that search the most for private health insurance in the first place. Moreover, individuals that smoke more are associated with a positive effect for their own health as well as to that of their family or community, and, conversely, reporting a worse general health state is associated with an increase at an individual and family level, but not clearly at a community level. Furthermore, high levels of alcoholic consumption also increase the probability of perceiving a high level of *Individual* and *Family Risk*, but not of *Community Risk*. These two results might be a sign that people do not consider their individual habits relevant enough to influence their community's health, which could be hazardous when public health rests all its citizens' individual actions.

Concerning demographical characteristics, the results point towards the fact that women, older individuals and members of smaller households are associated with a higher probability of perceiving a high risk. The former seems to contradict the ideas expressed by Bruine de Bruin

(2020), that the elderly might be more prone to being optimistic in order to maximize their psychological well-being during their incredibly stressing period, and instead confirms the *à priori* conception that, as they are more vulnerable, they should perceive a higher risk level. The latter contradicts what would be expected since the usual idea is that minimizing the number of close contacts helps containing the virus but data points towards this leading to an increase of the chance of someone having a high risk perception, possibly due to some notion of being a constrict group or attaining group immunity within their household. Furthermore, higher levels of education seem to lead people to be more likely to highly fear the new Coronavirus, possibly due to some heightened attention to the uncertainty of the scientific knowledge surrounding the pandemic or to a notion of the being able to have more precise information on the fact.

Risk perceptions, as posited by PMT, provide the necessary information for a person to conduct their own cost-benefit analysis of taking preventive action to minimize the damaging effects of the pandemic. A higher risk perception means that individuals feel more vulnerable to the pandemic, i.e. increasing the costs of not implementing an adaptive strategy and making people more likely to do so. Therefore, taking this information into account is extremely important as it helps target certain groups of individuals in order to guarantee that they perceive the correct risk level and cooperate with state authorities. Thus, targeting the young or people with low levels of education, which are associated with a higher chance of perceiving a low risk-level, for example, could help them adjust their expectations and avoid underestimating the real level of risk in order to better minimize the hazardous aspects of the current crisis.

Further research could be done to estimate the accurate level of risk people should have in mind regarding COVID-19. In terms of extending this project's approach, it could be helpful to have data on whether someone has been subject to COVID-19 either on an individual or collective manner, and to attempt to better understand if risk perceptions influence preventive behaviours or if it is the other way around, in this specific context. Moreover, the GFK2020 data was collected quite early in the pandemic crisis, a time in which, for example, not many pre-existing conditions had been identified as constituting risk-factors for one's vulnerability to the COVID-19. Thus, it could be interesting to see if there was some change in the importance of these factors due to the uncover of new information regarding its importance and worsening effect of the consequences deriving from an infection by the new Coronavirus. Finally, a randomized control trial could also be useful in better comprehending inner mechanisms of this notion's formulation associated with the new Coronavirus, so as to circumvent the fact that this paper uses reported data and not information collected in an objectively comparable manner.

7. Limitations

Nonetheless, this project is far from perfect and there are some points that were addressed during the methodological formulation that should be taken into consideration. First of all, as data had already been collected, it was not possible to include certain variables suggested in past literature, such as cognitive and socio-cultural parameters as suggested in Dryhurst et al. 2020 or elements related to the evolution of the pandemic like those used in Breznau 2020.

Also concerning the available data, the usage of so many variables and so many categorical variables may have damaged the precision of the regressions. Nonetheless, in such a unique time, when data is scarce, being able to get the faintest idea regarding the determinants of risk perceptions is already a positive and valuable contribution for the scientific community, for policymakers, and for society as a whole.

Furthermore, this paper initially intended to use a trivariate MLE model to test whether a system composed by the three Probit equations would be a better fit than the SUR model for the data, which was not possible due to the great computational power necessary to run such a model.

In terms of the analysis of the results, although looking at these general tendencies is a clear and easy method for the purpose of aiding policymakers to best target their measures, the interpretation of what these associations might mean to adaptive behaviours is not as direct or easy. In fact, there is the possibility of some reverse causality, as risk perceptions might be what leads people to protect themselves and prevent being infected with the new Coronavirus, or the other way around, with people having a given risk perceptions because of the coping mechanisms they have put in place, or lack thereof. Nonetheless, in this paper, it has been chosen to use the PMT to surpass this setback to be able to inform on such a crucial topic to the fight against COVID-19.

8. References

Breznau, Nate. 2020. "The Welfare State and Risk Perceptions: The Novel Coronavirus Pandemic and Public Concern in 70 Countries." European Societies, August, 1–14. https://doi.org/10.1080/14616696.2020.1793215.

Bruine de Bruin, Wändi. 2020. "Age Differences in COVID-19 Risk Perceptions and Mental Health: Evidence From a National U.S. Survey Conducted in March 2020." Edited by Derek Isaacowitz. The Journals of Gerontology: Series B XX (XX). https://doi.org/10.1093/geronb/gbaa074.

Carstensen, L. L. 2006. "The Influence of a Sense of Time on Human Development." Science 312 (5782): 1913–15. https://doi.org/10.1126/science.1127488.

Charles, Susan Turk. 2010. "Strength and Vulnerability Integration: A Model of Emotional Well-Being across Adulthood." Psychological Bulletin 136 (6): 1068–91. https://doi.org/10.1037/a0021232.

Cook, D. G., and S. J. Pocock. 1983. "Multiple Regression in Geographical Mortality Studies, with Allowance for Spatially Correlated Errors." Biometrics 39 (2): 361–371. https://doi.org/10.2307/2531009.

Dryhurst, Sarah, Claudia R. Schneider, John Kerr, Alexandra L. J. Freeman, Gabriel Recchia, Anne Marthe van der Bles, David Spiegelhalter, and Sander van der Linden. 2020. "Risk Perceptions of COVID-19 around the World." Journal of Risk Research, May, 1–13. https://doi.org/10.1080/13669877.2020.1758193.

Floyd, Donna L., Steven Prentice-Dunn, and Roland W. Rogers. 2000. "A Meta-Analysis of Research on Protection Motivation Theory." Journal of Applied Social Psychology 30 (2): 407–29. https://doi.org/10.1111/j.1559-1816.2000.tb02323.x.

McFadden, SarahAnn M., Amyn A. Malik, Obianuju G. Aguolu, Kathryn S. Willebrand, and Saad B. Omer. 2020. "Perceptions of the Adult US Population Regarding the Novel Coronavirus Outbreak." Edited by Abdallah M. Samy. PLOS ONE 15 (4): e0231808. https://doi.org/10.1371/journal.pone.0231808.

World Health Organization. n.d. "Timeline: WHO's COVID-19 Response." Www.Who.Int. Accessed November 25, 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline?gclid=CjwKCAiAnvj9BRA4EiwAuUMDf7mlJMFKoDsvTBGPUiu1BAiYGyonKg0r0t3XoLRe6bgaTjwHD8uNKxoCFJcQAvD_BwE#event-1.

Annex I – General Appendix

Region GFK Metris	Interviews	Percentage
Northern Seaside	265	20.8
Great Oporto	156	12.3
Interior	168	13.2
Central Seaside	219	17.2
Great Lisbon	360	28.3
Alentejo	52	4.1
Algarve	51	4.0
Total	1,271	

Table 3 – Geographical distribution of the sample per Region

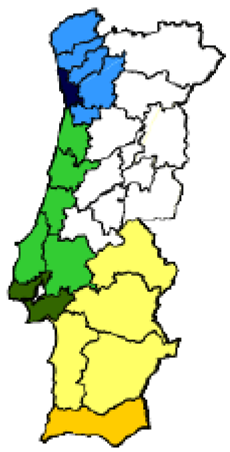

Figure 1 – Geographical distribution of the defined Regions in a map according to the colour scheme in Table 1

Individual Risk Binary Form[13]	LPM	MLE's ME	SUR
Average Daily % Change in the Number of COVID Cases	1.968	2.158	1.968
	(0.584)	(0.632)	(0.574)
Asthma	0.145	0.239*	0.145
	(0.125)	(0.100)	(0.115)
Lung Disease	0.039	0.067	0.039
	(0.708)	(0.661)	(0.701)
Heart Attack	0.147	0.187	0.147
	(0.309)	(0.455)	(0.296)
Heart Disease	0.031	0.033	0.031
	(0.775)	(0.835)	(0.769)
Hypertension	0.060	0.100	0.060
	(0.245)	(0.190)	(0.233)
Stroke	0.010	-0.099	0.010
	(0.946)	(0.709)	(0.944)
Arthritis	0.023	0.120	0.023
	(0.774)	(0.400)	(0.768)
Back Pain	-0.008	0.053	-0.008
	(0.930)	(0.723)	(0.928)
Neck Pain	0.030	0.042	0.030
	(0.804)	(0.823)	(0.799)
Diabetes	0.174***	0.352***	0.174***
	(0.001)	(0.000)	(0.001)
Allergies	0.308**	0.394**	0.308**
	(0.013)	(0.014)	(0.011)
Hepatic Cirrhosis	-0.099	-0.112	-0.099
	(0.574)	(0.672)	(0.564)
Bladder Control	-0.048	0.000	-0.048
	(0.822)	(.)	(0.817)
Kidney Disease	0.075	0.058	0.075
	(0.584)	(0.775)	(0.575)
Depression	-0.004	0.001	-0.004
	(0.958)	(0.995)	(0.957)
Cancer	0.106	0.087	0.106
	(0.422)	(0.662)	(0.410)
High Cholesterol	-0.034	-0.044	-0.034
	(0.896)	(0.910)	(0.893)
Private Health Insurance	0.007	0.015	0.007
	(0.860)	(0.764)	(0.857)
Smoke=1	0.165**	0.210**	0.165**
	(0.045)	(0.032)	(0.040)
Smoke=2	0.031	0.039	0.031

[13] Omitted Variables: Smoke Daily (*Smoke* = 0); Drink every day or almost every day (*Drink* = 0); Very Good General Health State (*General Health State* = 0); Ages between 15 and 24 years old (*Age by Interval* = 0); Households with up to 2 members (*Household Members by Interval* = 0); University/ Post-Graduate/ Master's/ Doctorate Education level (*Education* = 0); Self-Employed (*Occupation* = 0); Northern Seaside Region (*Region* = 0).

	(0.538)	(0.546)	(0.527)
Smoke=3	0.003	-0.005	0.003
	(0.940)	(0.912)	(0.939)
Drink=1	0.004	0.015	0.004
	(0.949)	(0.870)	(0.948)
Drink=2	-0.020	-0.018	-0.020
	(0.700)	(0.792)	(0.692)
Drink=3	-0.121**	-0.157**	-0.121**
	(0.017)	(0.013)	(0.015)
Drink=4	-0.039	-0.033	-0.039
	(0.435)	(0.600)	(0.422)
Drink=5	-0.039	-0.046	-0.039
	(0.613)	(0.646)	(0.604)
Drink=6	-0.042	-0.052	-0.042
	(0.404)	(0.428)	(0.392)
Drink=7	-0.010	0.000	-0.010
	(0.883)	(0.997)	(0.880)
Drink=8	0.015	0.025	0.015
	(0.739)	(0.680)	(0.732)
General Health State=1	0.025	0.022	0.025
	(0.506)	(0.623)	(0.495)
General Health State=2	0.239***	0.244***	0.239***
	(0.000)	(0.000)	(0.000)
General Health State=3	0.371***	0.560***	0.371***
	(0.000)	(0.000)	(0.000)
General Health State=4	0.440	0.000	0.440
	(0.117)	(.)	(0.107)
Gender	0.033	0.046	0.033
	(0.242)	(0.215)	(0.230)
Age by Interval=1	0.116**	0.142**	0.116**
	(0.045)	(0.045)	(0.040)
Age by Interval=2	0.125**	0.158**	0.125**
	(0.030)	(0.023)	(0.026)
Age by Interval=3	0.105*	0.127*	0.105*
	(0.086)	(0.085)	(0.078)
Age by Interval=4	0.194***	0.219***	0.194***
	(0.004)	(0.007)	(0.003)
Age by Interval=5	0.284***	0.376***	0.284***
	(0.001)	(0.001)	(0.001)
Age by Interval=6	0.274***	0.338***	0.274***
	(0.004)	(0.006)	(0.003)
Household Members by Interval=1	0.006	0.014	0.006
	(0.826)	(0.718)	(0.822)
Household Members by Interval=2	-0.029	-0.034	-0.029
	(0.570)	(0.610)	(0.560)
Education=1	0.178	0.214	0.178
	(0.140)	(0.133)	(0.130)
Education=2	0.190*	0.239**	0.190**

	(0.054)	(0.044)	(0.048)
Education=3	0.010	0.010	0.010
	(0.843)	(0.871)	(0.839)
Education=4	0.065	0.088	0.065
	(0.207)	(0.180)	(0.196)
Education=5	0.057	0.063	0.057
	(0.329)	(0.395)	(0.316)
Education=6	0.078	0.104	0.078
	(0.187)	(0.171)	(0.176)
Education=7	0.075	0.073	0.075
	(0.388)	(0.574)	(0.375)
Occupation=1	0.005	-0.002	0.005
	(0.911)	(0.978)	(0.909)
Occupation=2	-0.044	-0.081	-0.044
	(0.434)	(0.248)	(0.422)
Occupation=3	0.056	0.049	0.056
	(0.409)	(0.571)	(0.397)
Occupation=4	0.042	0.023	0.042
	(0.604)	(0.830)	(0.595)
Occupation=5	0.007	0.002	0.007
	(0.932)	(0.989)	(0.931)
Region=1	-0.085*	-0.113*	-0.085*
	(0.066)	(0.057)	(0.059)
Region=2	-0.119***	-0.165***	-0.119***
	(0.008)	(0.005)	(0.007)
Region=3	-0.113***	-0.144***	-0.113***
	(0.006)	(0.007)	(0.005)
Region=4	-0.027	-0.031	-0.027
	(0.747)	(0.764)	(0.741)
Region=5	-0.020	-0.037	-0.020
	(0.794)	(0.714)	(0.788)
Region=6	-0.289***	-0.396***	-0.289***
	(0.000)	(0.000)	(0.000)
Constant	0.158*	0.000	0.158*
	(0.096)	(.)	(0.087)
Observations	1,230	1,222	1,230
R^2	0.285		0.285
Adjusted R^2	0.248	0.155	0.248
Pseudo R^2		0.239	

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4 – Estimates for the Individual Risk in Binary Form under the LPM, MLE and SUR approaches

Family Risk Binary Form[14]	LPM	MLE's ME	SUR
Average Daily % Change in the Number of COVID Cases	0.603	0.120	0.603
	(0.868)	(0.978)	(0.865)
Asthma	0.152	0.264*	0.152
	(0.114)	(0.066)	(0.104)
Lung Disease	0.115	0.182	0.115
	(0.274)	(0.241)	(0.261)
Heart Attack	0.049	0.030	0.049
	(0.736)	(0.882)	(0.730)
Heart Disease	0.107	0.167	0.107
	(0.337)	(0.301)	(0.325)
Hypertension	0.033	0.049	0.033
	(0.524)	(0.472)	(0.513)
Stroke	-0.026	-0.041	-0.026
	(0.865)	(0.871)	(0.862)
Arthritis	-0.118	-0.163	-0.118
	(0.153)	(0.141)	(0.142)
Back Pain	0.090	0.173	0.090
	(0.334)	(0.204)	(0.321)
Neck Pain	-0.142	-0.268*	-0.142
	(0.246)	(0.092)	(0.233)
Diabetes	0.131**	0.211***	0.131**
	(0.016)	(0.005)	(0.013)
Allergies	0.477***	0.647***	0.477***
	(0.000)	(0.000)	(0.000)
Hepatic Cirrhosis	-0.052	-0.080	-0.052
	(0.770)	(0.752)	(0.764)
Bladder Control	0.106	0.000	0.106
	(0.624)	(.)	(0.615)
Kidney Disease	0.098	0.052	0.098
	(0.482)	(0.771)	(0.471)
Depression	-0.021	-0.019	-0.021
	(0.798)	(0.851)	(0.793)
Cancer	0.030	0.003	0.030
	(0.822)	(0.985)	(0.817)
High Cholesterol	0.015	0.039	0.015
	(0.955)	(0.912)	(0.953)
Private Health Insurance	0.005	0.010	0.005
	(0.901)	(0.844)	(0.898)
Smoke=1	0.009	0.035	0.009
	(0.919)	(0.733)	(0.916)
Smoke=2	0.019	0.023	0.019

[14] Omitted Variables: Smoke Daily (*Smoke* = 0); Drink every day or almost every day (*Drink* = 0); Very Good General Health State (*General Health State* = 0); Ages between 15 and 24 years old (*Age by Interval* = 0); Households with up to 2 members (*Household Members by Interval* = 0); University/ Post-Graduate/ Master's/ Doctorate Education level (*Education* = 0); Self-Employed (*Occupation* = 0); Student (*Occupation* = 5; Northern Seaside Region (*Region* = 0).

	(0.702)	(0.714)	(0.694)
Smoke=3	0.003	-0.004	0.003
	(0.942)	(0.934)	(0.940)
Drink=1	0.059	0.074	0.059
	(0.403)	(0.408)	(0.391)
Drink=2	-0.004	-0.004	-0.004
	(0.946)	(0.947)	(0.944)
Drink=3	-0.050	-0.064	-0.050
	(0.330)	(0.308)	(0.317)
Drink=4	0.048	0.072	0.048
	(0.346)	(0.252)	(0.333)
Drink=5	-0.048	-0.058	-0.048
	(0.538)	(0.561)	(0.528)
Drink=6	-0.060	-0.068	-0.060
	(0.244)	(0.284)	(0.232)
Drink=7	-0.019	-0.027	-0.019
	(0.775)	(0.752)	(0.769)
Drink=8	-0.014	-0.013	-0.014
	(0.759)	(0.821)	(0.753)
General Health State=1	-0.047	-0.069	-0.047
	(0.214)	(0.130)	(0.202)
General Health State=2	0.155^{***}	0.148^{**}	0.155^{***}
	(0.001)	(0.010)	(0.001)
General Health State=3	0.298^{***}	0.415^{***}	0.298^{***}
	(0.000)	(0.000)	(0.000)
General Health State=4	0.287	0.000	0.287
	(0.312)	(.)	(0.299)
Gender	0.053^{*}	0.074^{**}	0.053^{*}
	(0.063)	(0.040)	(0.057)
Age by Interval=1	0.119^{**}	0.147^{**}	0.119^{**}
	(0.043)	(0.031)	(0.037)
Age by Interval=2	0.155^{***}	0.195^{***}	0.155^{***}
	(0.008)	(0.004)	(0.006)
Age by Interval=3	0.102^{*}	0.136^{*}	0.102^{*}
	(0.100)	(0.056)	(0.091)
Age by Interval=4	0.207^{***}	0.237^{***}	0.207^{***}
	(0.002)	(0.003)	(0.002)
Age by Interval=5	0.258^{***}	0.318^{***}	0.258^{***}
	(0.004)	(0.004)	(0.003)
Age by Interval=6	0.254^{***}	0.290^{**}	0.254^{***}
	(0.008)	(0.016)	(0.006)
Household Members by Interval=1	-0.015	-0.017	-0.015
	(0.616)	(0.646)	(0.607)
Household Members by Interval=2	-0.021	-0.026	-0.021
	(0.681)	(0.687)	(0.673)
Education=1	0.214^{*}	0.249^{*}	0.214^{*}
	(0.081)	(0.084)	(0.073)
Education=2	0.121	0.153	0.121

	(0.225)	(0.216)	(0.213)
Education=3	0.036	0.042	0.036
	(0.459)	(0.461)	(0.447)
Education=4	0.113**	0.148**	0.113**
	(0.031)	(0.017)	(0.026)
Education=5	0.098*	0.111	0.098*
	(0.097)	(0.111)	(0.089)
Education=6	0.140**	0.181**	0.140**
	(0.020)	(0.012)	(0.017)
Education=7	0.058	0.032	0.058
	(0.508)	(0.765)	(0.497)
Occupation=1	-0.108**	-0.129**	-0.108**
	(0.017)	(0.017)	(0.014)
Occupation=2	-0.121**	-0.159**	-0.121**
	(0.035)	(0.020)	(0.030)
Occupation=3	0.004	-0.000	0.004
	(0.955)	(0.999)	(0.953)
Occupation=4	-0.048	-0.083	-0.048
	(0.559)	(0.416)	(0.549)
Occupation=5	-0.107	-0.132	-0.107
	(0.191)	(0.219)	(0.179)
Region=1	-0.010	-0.013	-0.010
	(0.836)	(0.827)	(0.832)
Region=2	-0.065	-0.095*	-0.065
	(0.157)	(0.098)	(0.147)
Region=3	-0.064	-0.082	-0.064
	(0.131)	(0.114)	(0.120)
Region=4	0.005	0.013	0.005
	(0.953)	(0.904)	(0.951)
Region=5	0.058	0.063	0.058
	(0.452)	(0.531)	(0.440)
Region=6	-0.258***	-0.314***	-0.258***
	(0.000)	(0.000)	(0.000)
Constant	0.227**	0.000	0.227**
	(0.019)	(.)	(0.016)
Observations	1,230	1,222	1,230
R^2	0.261		0.261
Adjusted R^2	0.222	0.127	0.222
Pseudo R^2		0.211	

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5 – Estimates for the Family Risk in Binary Form under the LPM, MLE and SUR approaches

Community Risk Binary Form[15]	LPM	MLE's ME	SUR
Average Daily % Change in the Number of COVID Cases	-3.082	-3.331	-3.082
	(0.446)	(0.438)	(0.422)
Asthma	0.106	0.128	0.106
	(0.287)	(0.301)	(0.294)
Lung Disease	-0.119	-0.182	-0.119
	(0.291)	(0.179)	(0.283)
Heart Attack	-0.047	-0.112	-0.047
	(0.741)	(0.563)	(0.761)
Heart Disease	0.007	-0.001	0.007
	(0.952)	(0.993)	(0.953)
Hypertension	0.014	0.006	0.014
	(0.798)	(0.933)	(0.805)
Stroke	0.071	0.200	0.071
	(0.484)	(0.468)	(0.659)
Arthritis	-0.068	-0.111	-0.068
	(0.405)	(0.312)	(0.436)
Back Pain	0.056	0.140	0.056
	(0.473)	(0.304)	(0.564)
Neck Pain	-0.073	-0.133	-0.073
	(0.542)	(0.405)	(0.572)
Diabetes	0.171***	0.242***	0.171***
	(0.001)	(0.001)	(0.003)
Allergies	0.216*	0.282*	0.216
	(0.077)	(0.089)	(0.103)
Hepatic Cirrhosis	-0.084	-0.141	-0.084
	(0.628)	(0.535)	(0.655)
Bladder Control	0.301**	0.000	0.301
	(0.016)	(.)	(0.188)
Kidney Disease	0.201*	0.352	0.201
	(0.055)	(0.152)	(0.170)
Depression	-0.039	-0.034	-0.039
	(0.654)	(0.736)	(0.649)
Cancer	-0.048	-0.094	-0.048
	(0.707)	(0.584)	(0.734)
High Cholesterol	0.029	0.027	0.029
	(0.873)	(0.938)	(0.917)
Private Health Insurance	0.031	0.034	0.031
	(0.466)	(0.470)	(0.453)
Smoke=1	0.017	0.028	0.017
	(0.856)	(0.773)	(0.845)
Smoke=2	-0.059	-0.066	-0.059

[15] Omitted Variables: Smoke Daily (*Smoke* = 0); Drink every day or almost every day (*Drink* = 0); Very Good General Health State (*General Health State* = 0); Ages between 15 and 24 years old (*Age by Interval* = 0); Households with up to 2 members (*Household Members by Interval* = 0); University/ Post-Graduate/ Master's/ Doctorate Education level (*Education* = 0); Self-Employed (*Occupation* = 0); Student (*Occupation* = 5; Northern Seaside Region (*Region* = 0).

	(0.281)	(0.274)	(0.263)
Smoke=3	-0.043	-0.050	-0.043
	(0.282)	(0.233)	(0.253)
Drink=1	-0.070	-0.096	-0.070
	(0.332)	(0.254)	(0.344)
Drink=2	0.095*	0.111*	0.095*
	(0.093)	(0.083)	(0.091)
Drink=3	-0.038	-0.042	-0.038
	(0.503)	(0.495)	(0.485)
Drink=4	0.091	0.106*	0.091*
	(0.111)	(0.077)	(0.090)
Drink=5	0.088	0.100	0.088
	(0.301)	(0.277)	(0.280)
Drink=6	0.045	0.053	0.045
	(0.445)	(0.388)	(0.413)
Drink=7	0.081	0.093	0.081
	(0.262)	(0.250)	(0.251)
Drink=8	0.044	0.051	0.044
	(0.401)	(0.365)	(0.376)
General Health State=1	-0.030	-0.039	-0.030
	(0.475)	(0.377)	(0.455)
General Health State=2	-0.006	-0.018	-0.006
	(0.912)	(0.747)	(0.906)
General Health State=3	0.208***	0.311***	0.208**
	(0.005)	(0.001)	(0.019)
General Health State=4	-0.113	-0.122	-0.113
	(0.627)	(0.749)	(0.706)
Gender	0.016	0.022	0.016
	(0.606)	(0.519)	(0.591)
Age by Interval=1	0.124*	0.131*	0.124**
	(0.054)	(0.054)	(0.044)
Age by Interval=2	0.159**	0.172**	0.159***
	(0.015)	(0.010)	(0.009)
Age by Interval=3	0.128*	0.145**	0.128*
	(0.062)	(0.044)	(0.050)
Age by Interval=4	0.169**	0.188**	0.169**
	(0.023)	(0.018)	(0.019)
Age by Interval=5	0.265***	0.335***	0.265***
	(0.005)	(0.001)	(0.005)
Age by Interval=6	0.178*	0.182	0.178*
	(0.073)	(0.123)	(0.078)
Household Members by Interval=1	-0.015	-0.014	-0.015
	(0.649)	(0.703)	(0.631)
Household Members by Interval=2	0.012	0.011	0.012
	(0.830)	(0.855)	(0.826)
Education=1	0.030	0.034	0.030
	(0.820)	(0.811)	(0.814)
Education=2	0.113	0.128	0.113

	(0.270)	(0.287)	(0.282)
Education=3	0.045	0.053	0.045
	(0.402)	(0.362)	(0.389)
Education=4	0.118**	0.140**	0.118**
	(0.038)	(0.025)	(0.034)
Education=5	0.085	0.096	0.085
	(0.177)	(0.172)	(0.170)
Education=6	0.085	0.094	0.085
	(0.187)	(0.195)	(0.178)
Education=7	0.225***	0.310***	0.225**
	(0.005)	(0.004)	(0.015)
Occupation=1	-0.023	-0.030	-0.023
	(0.638)	(0.587)	(0.624)
Occupation=2	0.019	0.011	0.019
	(0.766)	(0.876)	(0.755)
Occupation=3	-0.000	-0.025	-0.000
	(0.997)	(0.766)	(0.997)
Occupation=4	0.075	0.088	0.075
	(0.369)	(0.385)	(0.379)
Occupation=5	0.130	0.130	0.130
	(0.148)	(0.166)	(0.131)
Region=1	0.006	0.009	0.006
	(0.910)	(0.872)	(0.905)
Region=2	-0.015	-0.025	-0.015
	(0.766)	(0.644)	(0.754)
Region=3	-0.143***	-0.158***	-0.143***
	(0.002)	(0.001)	(0.001)
Region=4	0.136	0.140	0.136
	(0.146)	(0.134)	(0.122)
Region=5	0.182**	0.184**	0.182**
	(0.025)	(0.033)	(0.026)
Region=6	-0.460***	-0.495***	-0.460***
	(0.000)	(0.000)	(0.000)
Constant	0.276**	0.000	0.276***
	(0.011)	(.)	(0.007)
Observations	1,230	1,225	1,230
R^2	0.153		0.153
Adjusted R^2	0.108	0.046	0.108
Pseudo R^2		0.128	

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6 - Estimates for the Community Risk in Binary Form under the LPM, MLE and SUR approaches

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance

Variables: fitted values of Individual Risk in Binary Form

chi2(1)	=	2.00
Prob > chi2	=	0.1573

Table 7 – Breusch-Pagan / Cook-Weisberg test for heteroskedasticity for the LPM regression of the binary form of Individual Risk Perception

White's test for heteroskedasticity
Ho: Homoskedasticity

Ha: Unrestricted Heteroskedasticity

chi2(979)	=	1006.19
Prob > chi2	=	0.2664

Table 8 – White's test for unrestricted heteroskedasticity for the LPM regression of the binary form of Individual Risk Perception

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance

Variables: fitted values of Family Risk in Binary Form

chi2(1)	=	0.01
Prob > chi2	=	0.9316

Table 9 – Breusch-Pagan / Cook-Weisberg test for heteroskedasticity for the LPM regression of the binary form of Family Risk Perception

White's test for heteroskedasticity
Ho: Homoskedasticity

Ha: Unrestricted Heteroskedasticity

chi2(979)	=	978.76
Prob > chi2	=	0.4961

Table 10 – White's test for unrestricted heteroskedasticity for the LPM regression of the binary form of Family Risk Perception

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance

Variables: fitted values of Community Risk in Binary Form

chi2(1) = 1.45
Prob > chi2 = 0.2284

Table 11 – Breusch-Pagan / Cook-Weisberg test for heteroskedasticity for the LPM regression of the binary form of Community Risk Perception

White's test for heteroskedasticity
Ho: Homoskedasticity

Ha: Unrestricted Heteroskedasticity

chi2(979) = 1077.59
Prob > chi2 = 0.0149

Table 12 – White's test for unrestricted heteroskedasticity for the LPM regression of the binary form of Community Risk Perception

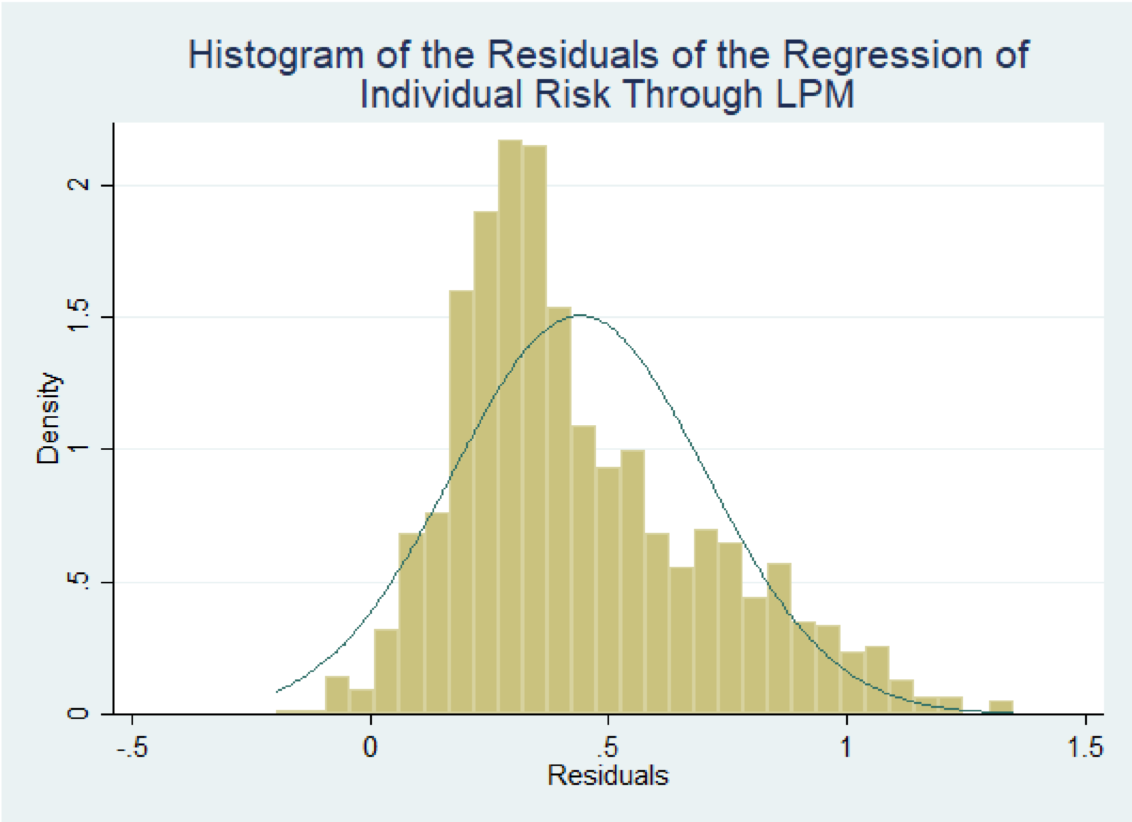

Figure 2 – Distribution of the residuals of the robust regression of Individual Risk in Binary Form, through LPM, plotted against the Normal Distribution

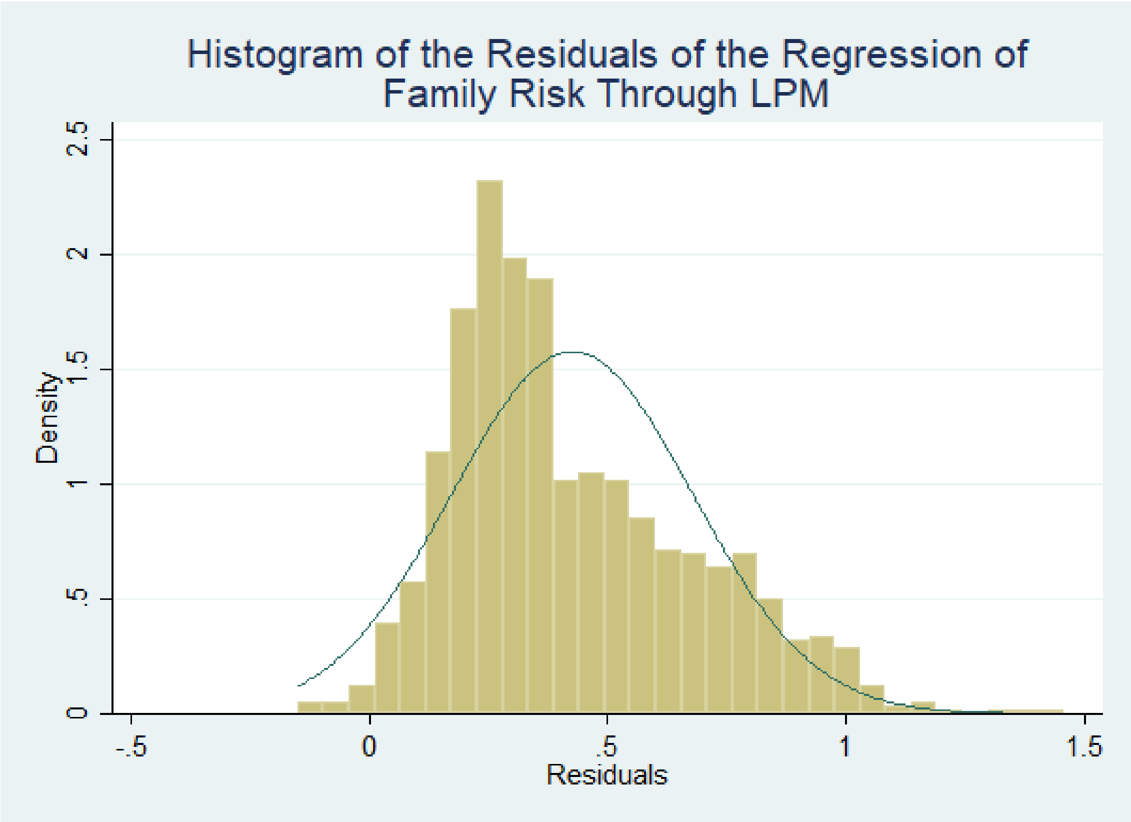

Figure 3 - Distribution of the residuals of the robust regression of Family Risk in Binary Form, through LPM, plotted against the Normal Distribution

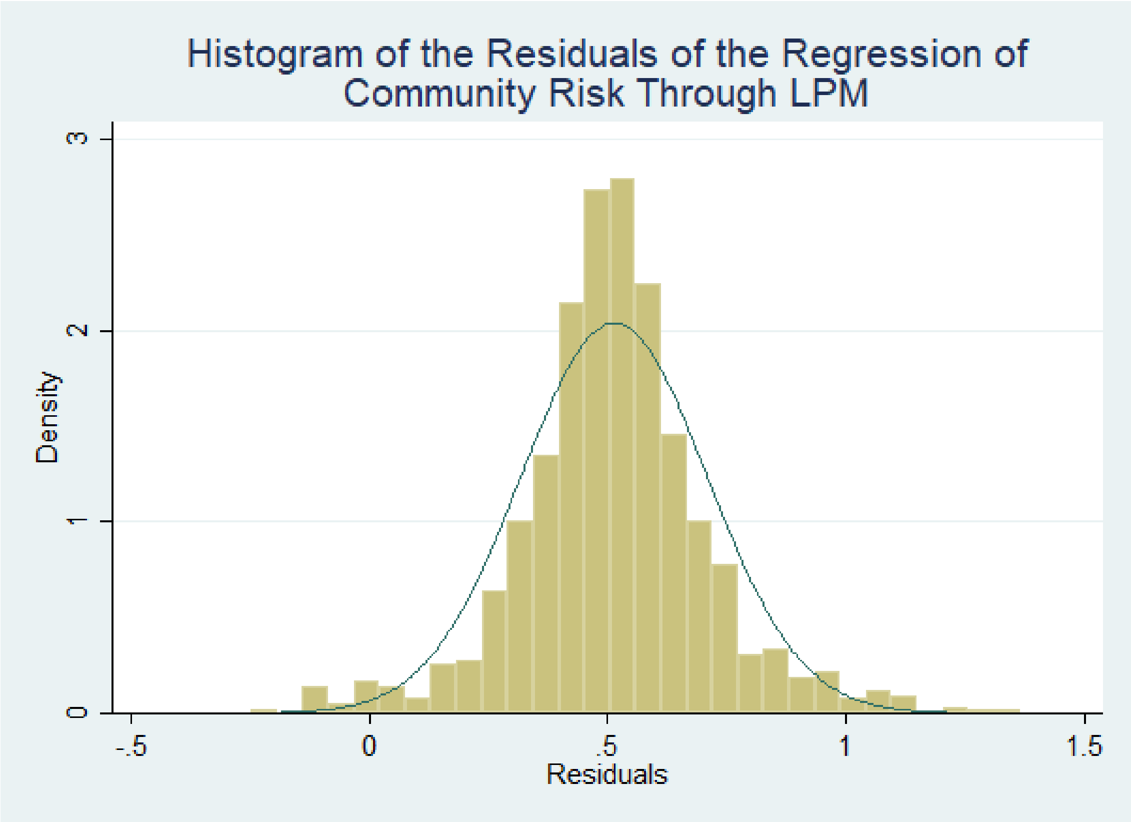

Figure 4 - Distribution of the residuals of the robust regression of Community Risk in Binary Form, through LPM, plotted against the Normal Distribution

Margins of Binary Forms	LPM & SUR	MLE
1.Individual Risk	0.438***	0.448***
	(0.000)	(0.000)
2.Family Risk	0.425***	0.419***
	(0.000)	(0.000)
3.Community Risk	0.513***	0.513***
	(0.000)	(0.000)
Observations of 1.	1,230	1,222
Observations of 2.	1,230	1,222
Observations of 3.	1,230	1,225

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13 – Margins for the three dependent variables at the covariates' mean values

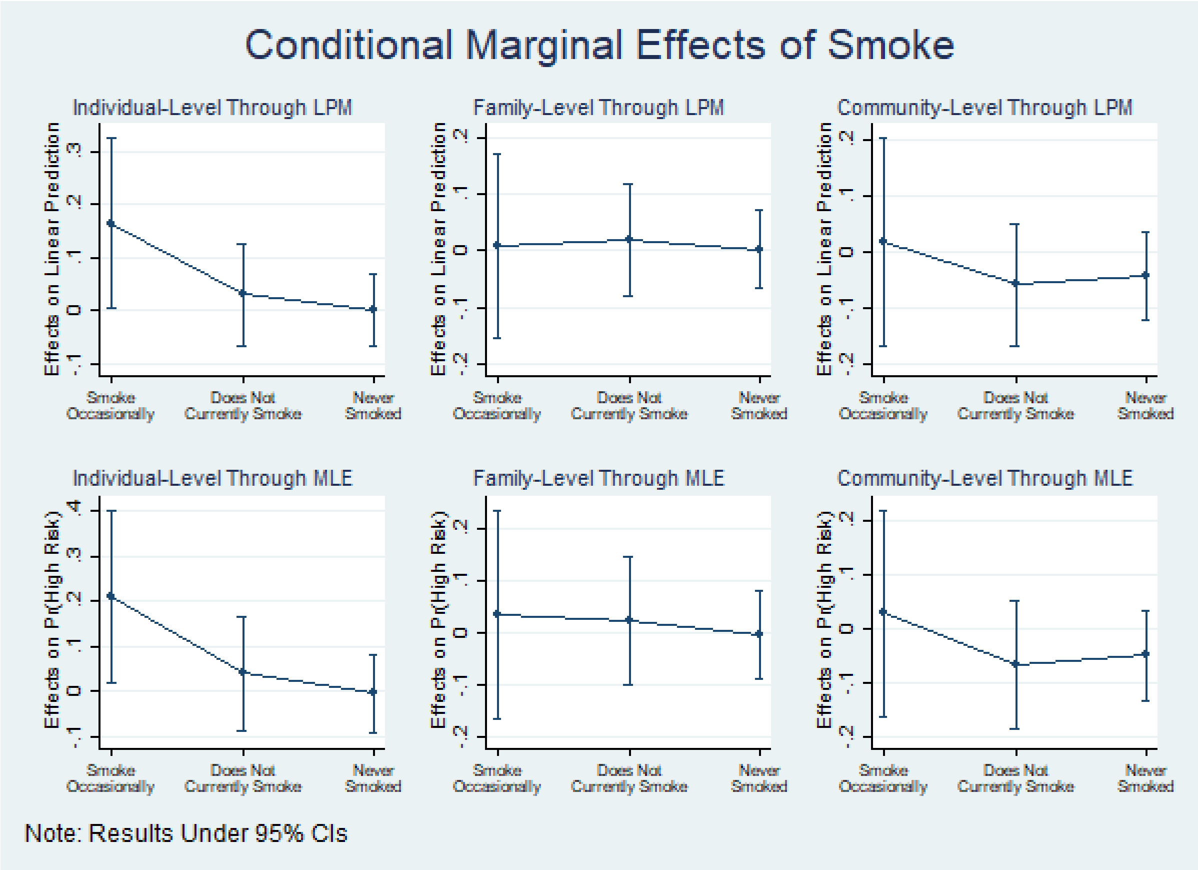

Figure 5 – Conditional Marginal Effects of the variable Smoke on all three levels of proximity and through all three methods

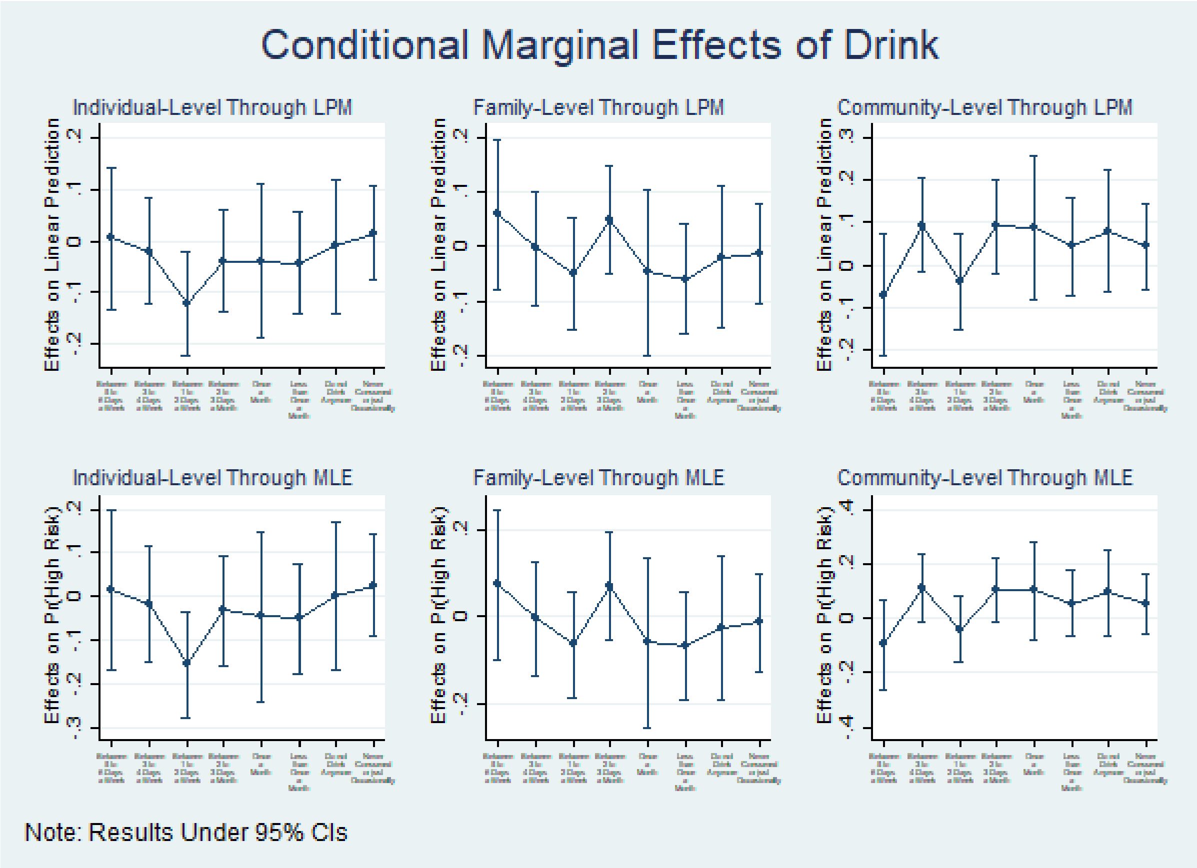

Figure 6 – Conditional Marginal Effects of the variable Drink on all three levels of proximity and through all three methods

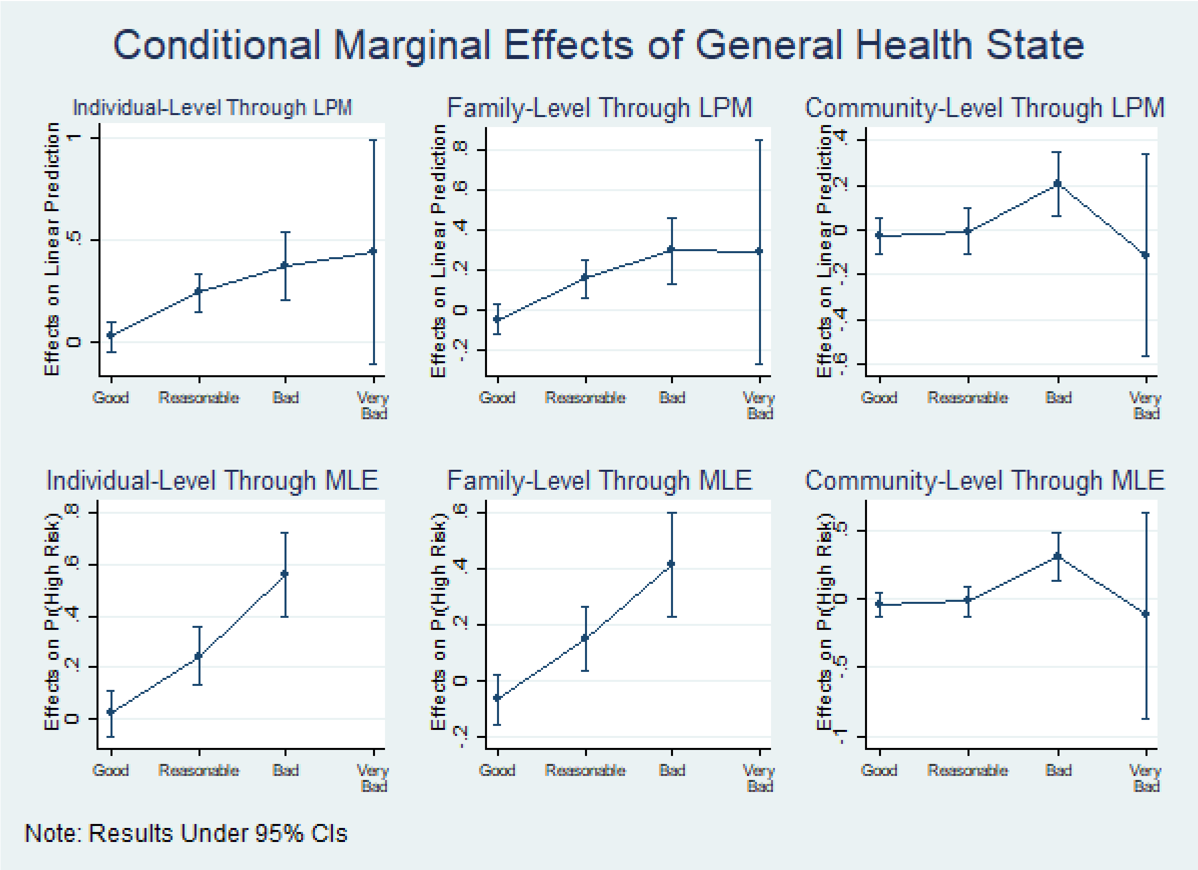

Figure 7 – Conditional Marginal Effects of the variable General Health State on all three levels of proximity and through all three methods

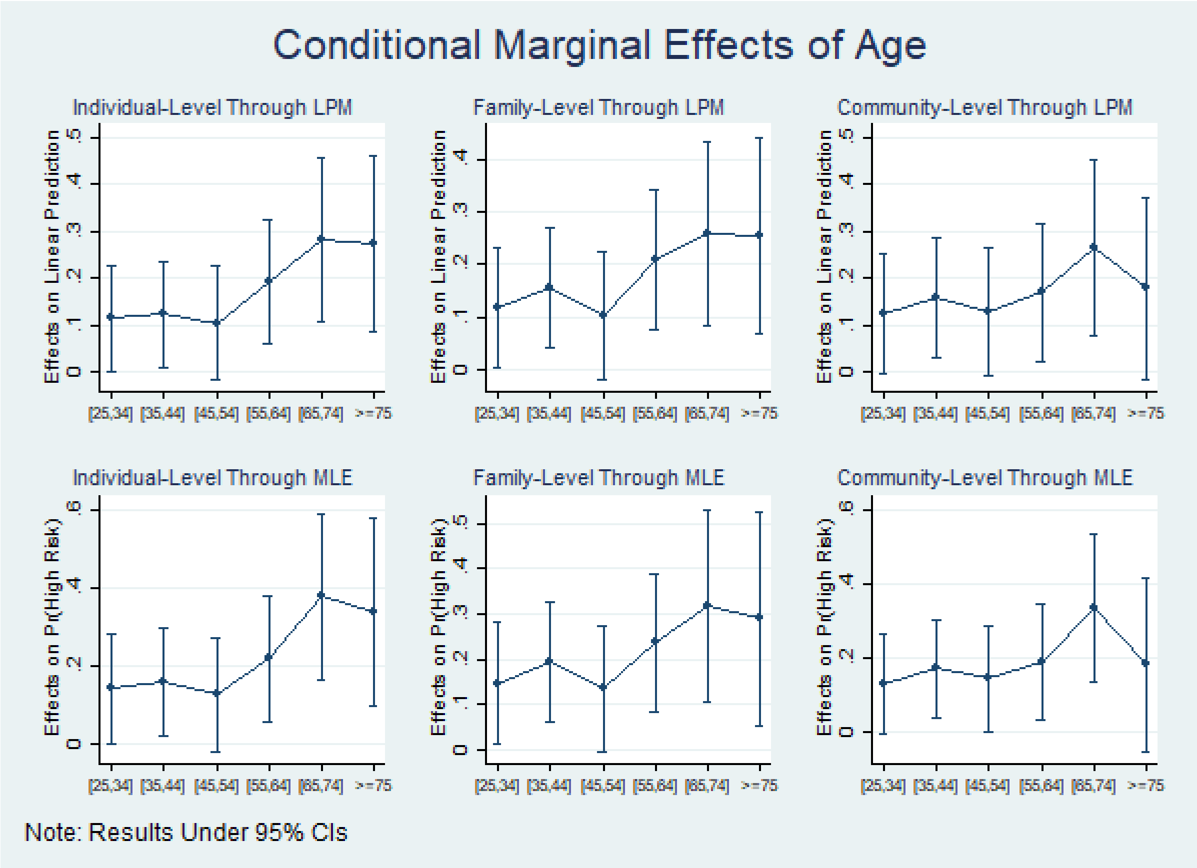

Figure 8 – Conditional Marginal Effects of the variable Age on all three levels of proximity and through all three methods

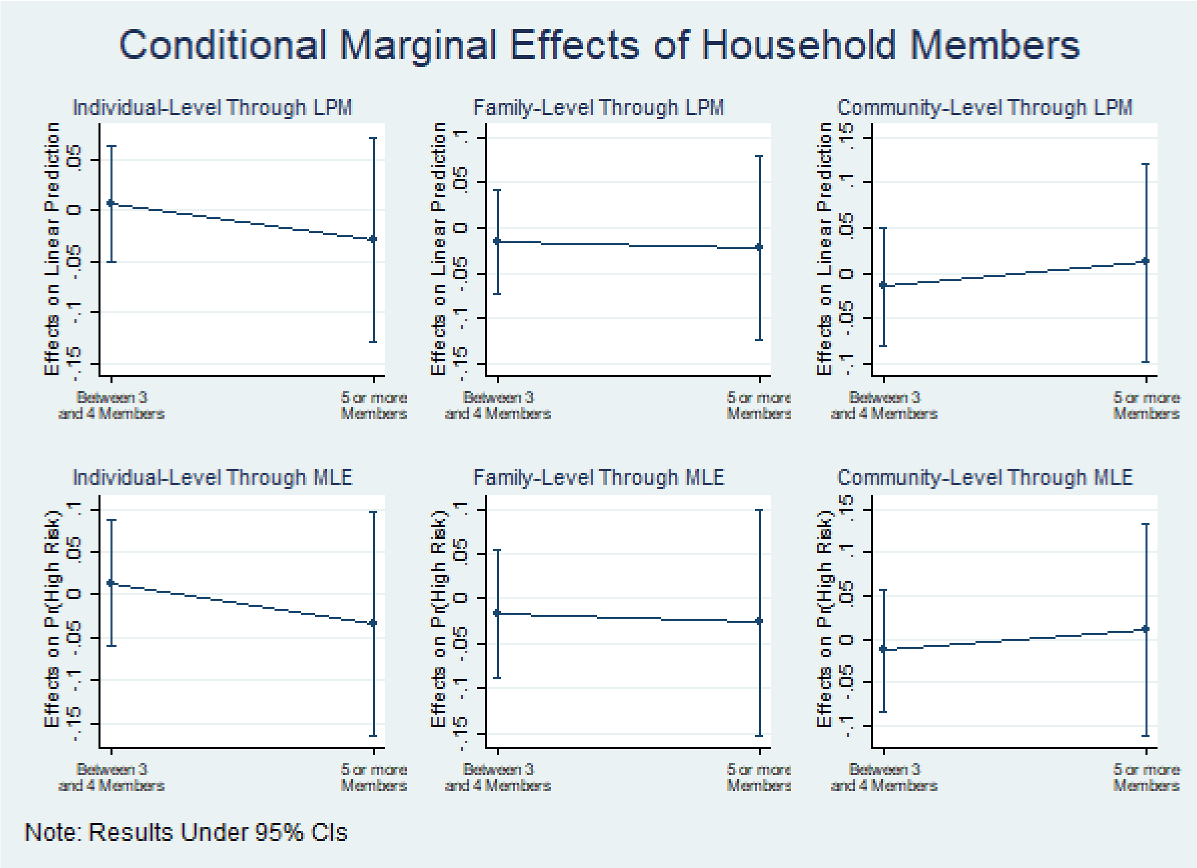

Figure 9 – Conditional Marginal Effects of the variable Household Members on all three levels of proximity and through all three methods

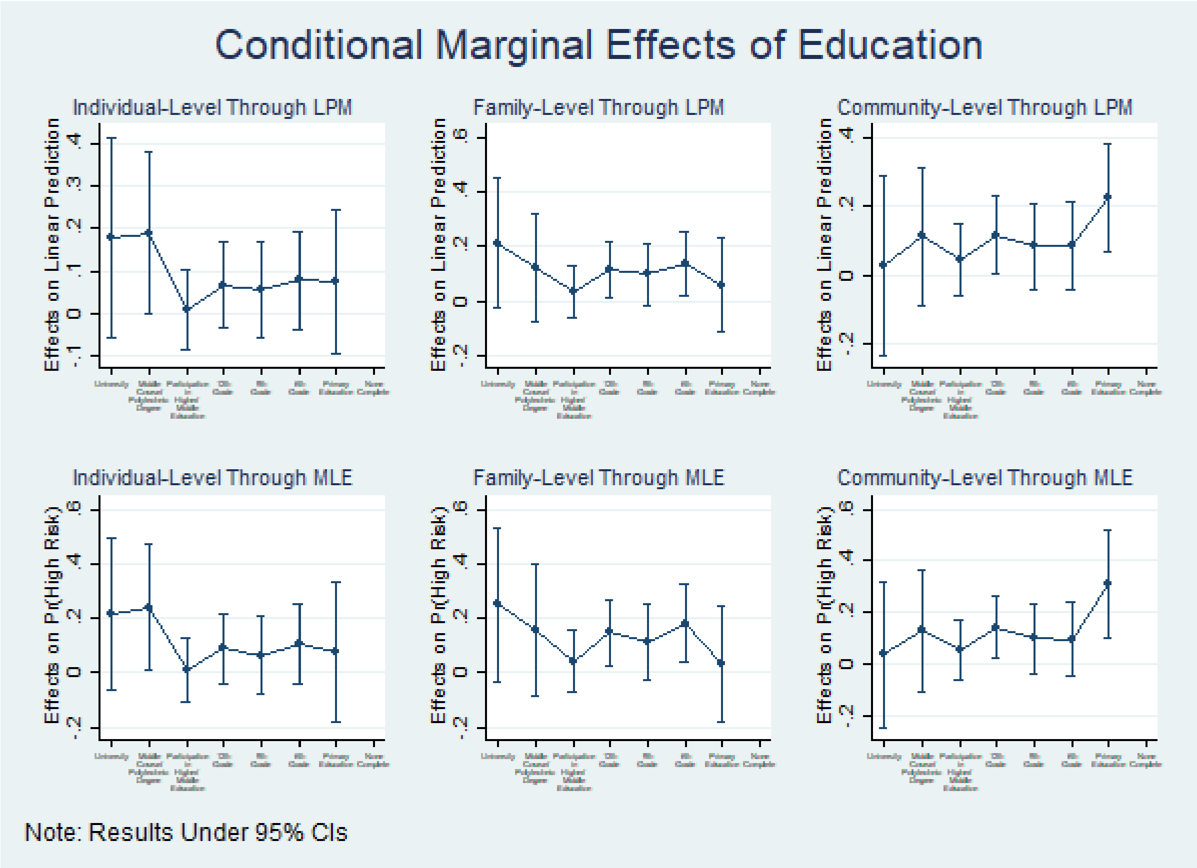

Figure 10 - Conditional Marginal Effects of the variable Education on all three levels of proximity and through all three methods

Annex II – Variables

Variable	Question	Obs	Mean	Std.Dev.
Survey Number		1,271	500,000	5,448.532
Average Daily % Change in the Number of COVID Cases[16]	The arithmetic average of the percentage change in the cumulative number of cases per region in the week prior to the survey.[17][18]	1,271	.008	.009
Pre-Existing Conditions	"Do you have any chronic condition that requires specific medication?"	1,271	.752	.432
Asthma	"Could you tell me which disease(s) is/are that/those?"	1,271	.02	.142
Lung Disease		1,271	.017	.13
Heart Attack		1,271	.008	.088
Heart Disease		1,271	.014	.118
Hypertension		1,271	.092	.289
Stroke		1,271	.008	.088
Arthritis		1,271	.032	.177
Back Pain		1,271	.025	.157
Neck Pain		1,271	.013	.112
Diabetes		1,271	.079	.269
Allergies		1,271	.01	.101
Hepatic Cirrhosis		1,271	.006	.074
Bladder Control		1,271	.004	.063
Kidney Disease		1,271	.009	.093
Depression		1,271	.027	.161
Cancer		1,271	.009	.097
High Cholesterol		1,271	.002	.049
Hand Washing	"The COVID-19 pandemic altered, in the past few months, our daily lives. Have you adapted any of the following behaviours in order to prevent contagion?" • Wash my hands frequently • Use mask when I go out • Do not shake hands • Try to maintain social distancing • Avoid entering crowded places • Try to go to the supermarket as little as possible • Stopped meeting with friends • Stopped meeting with family that does not live with me • Do not leave the house unless it is an emergency • None of the above	1,271	.949	.22
Mask		1,271	.844	.363
No Handshaking		1,271	.833	.373
Social Distancing		1,271	.876	.33
No Crowds		1,271	.791	.407
No Supermarkets		1,271	.673	.469
No Friends		1,271	.601	.49
No Family		1,271	.552	.497
No Outside		1,271	.375	.484
No Behaviour		1,271	.008	.088

[16] This is the only variable not taken from the GFK2020. Due to an overlap of the Interior region (as expressed in the GFK2020) with both the Regional Health Authority (RHA) of the North and of the Centre, as expressed in the data format adopted by the Portuguese National Health Authority, it was necessary to match the regions in this data with those in the GFK2020, as shown in Table 53 in Annex II. More precisely, as the values for the RHA of the North were higher than those of the Centre, the author opted for assigning the values reported by the RHA North to the whole region of the interior, deeming it better to overestimate that entire region rather than underestimating it.

[17] See Equation 1 in Annex III

[18] In this paper, the transformation is performed using an arithmetic average (instead of, for example, a geometric average) due to its simplicity and the fact than each average only contained 7 days at a time, which meant that the difference in the final results included in the regression would not be expressive enough to change the final outcome.

General Risk	“We would like to understand what you think and feel regarding the risk associated with the spreading of COVID-19. Please, use a scale to evaluate your probability of being infected with the new Coronavirus and its possible consequences. To do so, use a scale from 1 to 5 in which ‘1’ means ‘No risk’ and ‘5’ means ‘Very high risk’. You can use any intermediate value to provide us with your evaluation.”: • “My risk of getting infected with the Coronavirus” (General Risk) • “The risk for my health due to COVID-19” (Individual Risk) • “The risk for the health of my family’s members due to COVID-19” (Family Risk) • “The risk for the health of my community’s members due to COVID-19” (Community Risk)	1,224	3.258	1.082
Individual Risk		1,228	3.305	1.115
Family Risk		1,223	3.455	1.027
Community Risk		1,198	3.453	.975
Private Health Insurance	“Do you have any private health insurance?”	1,268	.851	.356
Smoke	“Which of the following statements best describes your situation regarding tobacco consumption? Would you say you…”	1,249	1.908	1.359
Drink	“Which of the following statements best describes your situation regarding the consumption of alcoholic beverages? Would you say that you consume alcoholic beverages…”	1,242	3.907	3.019
General Health State	“In general, how would you consider the state of your health? Would you that it is…”	1,271	1.194	.782
Gender	“Register Gender”	1,271	.526	.5
Age	“Please, tell me, how old are you?”	1,271	45.487	17.815
Household Members	“How many people live in your house, including yourself, both adults and children?”	1,271	2.703	1.239
Education	“And what is the highest level of instruction you have completed?”	1,271	3.886	1.789
Occupation	“What is your current occupation/profession?”	1,271	1.821	1.374
Income	“Remembering that responses are totally anonymous and will be treated in an aggregated form, we would like to count with your collaboration for one last characterization question. In a typical month, in which would you place your Net Income? Consider any regular income source: wages, social transfers (pensions, child support, subsidies, etc.), income derived from capital (interest on bank deposits and bonds, share dividends, etc.), rents, transfers from other households (alimony and other regular transfers).”	646	3.379	2.155
District	Registered by the Interviewer	1,271	9.242	4.997
Municipality	Registered by the Interviewer	1,271	131.872	76.18
Region	Registered by the Interviewer	1,271	2.482	1.744
Age by Interval	Categorical version of the variable Age comprised of 7 brackets.	1,271	2.574	1.756
Household Members by	Categorical version of the variable	1,271	.598	.627

Interval	Household Members comprised of 3 brackets.			

Table 14 – Descriptive statistics of all variables taken under consideration for inclusion in the models

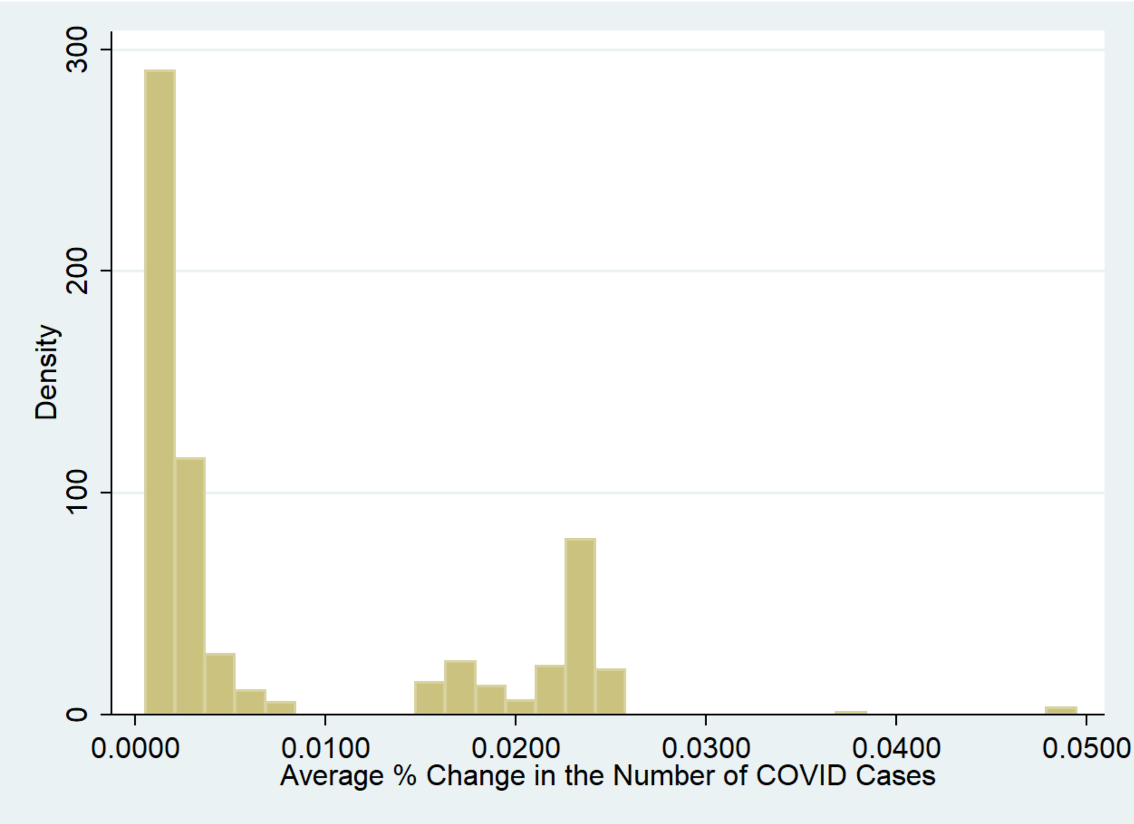

Figure 11 - Histogram of the variable AVG_CASES

Average Daily % Change in the Number of COVID Cases	Freq.	Percent	Cum.
.0005211	13	1.02	1.02
.0006418	34	2.68	3.70
.000701	18	1.42	5.11
.0007854	52	4.09	9.21
.000861	13	1.02	10.23
.0008617	41	3.23	13.45
.0009286	42	3.30	16.76
.0009872	23	1.81	18.57
.0011076	41	3.23	21.79
.0012656	37	2.91	24.70
.001326	1	0.08	24.78
.0013568	1	0.08	24.86
.001363	1	0.08	24.94
.0014219	1	0.08	25.02
.0014246	2	0.16	25.18
.0014589	15	1.18	26.36

.0014671	15	1.18	27.54
.0015081	10	0.79	28.32
.0015314	16	1.26	29.58
.0015573	4	0.31	29.90
.0015895	7	0.55	30.45
.0015951	22	1.73	32.18
.0015988	1	0.08	32.26
.0016049	23	1.81	34.07
.0016748	10	0.79	34.85
.0016801	16	1.26	36.11
.0016917	15	1.18	37.29
.0017698	8	0.63	37.92
.0017792	3	0.24	38.16
.0018537	22	1.73	39.89
.0018537	3	0.24	40.13
.0018637	11	0.87	40.99
.0018965	7	0.55	41.54
.0019411	45	3.54	45.08
.0019986	10	0.79	45.87
.0021372	8	0.63	46.50
.0022213	2	0.16	46.66
.0022954	3	0.24	46.89
.0023222	27	2.12	49.02
.0023515	7	0.55	49.57
.0023624	6	0.47	50.04
.0024232	8	0.63	50.67
.0024325	5	0.39	51.06
.0024482	4	0.31	51.38
.0024659	20	1.57	52.95
.0024779	6	0.47	53.42
.0024788	33	2.60	56.02
.0026013	1	0.08	56.10
.0026204	31	2.44	58.54
.0028072	6	0.47	59.01
.0028249	3	0.24	59.24
.00288	10	0.79	60.03
.0029892	8	0.63	60.66
.0031688	9	0.71	61.37
.0031692	4	0.31	61.68
.0033191	1	0.08	61.76
.0033593	10	0.79	62.55
.003425	13	1.02	63.57
.0035109	3	0.24	63.81
.0035302	3	0.24	64.04
.0035817	1	0.08	64.12
.0038664	3	0.24	64.36
.0039786	15	1.18	65.54
.0040016	3	0.24	65.77
.004083	2	0.16	65.93
.004098	1	0.08	66.01
.0042607	11	0.87	66.88
.0043024	4	0.31	67.19
.0043385	1	0.08	67.27
.0044976	3	0.24	67.51
.0048994	5	0.39	67.90
.005047	3	0.24	68.14
.0051757	4	0.31	68.45
.0053861	3	0.24	68.69
.0056239	2	0.16	68.84
.0057477	4	0.31	69.16

.0057869	1	0.08	69.24
.0063772	1	0.08	69.32
.0064255	3	0.24	69.55
.0064269	8	0.63	70.18
.0069699	4	0.31	70.50
.0071977	1	0.08	70.57
.0078897	5	0.39	70.97
.0080016	1	0.08	71.05
.0150636	11	0.87	71.91
.0153445	8	0.63	72.54
.01553	2	0.16	72.70
.016162	2	0.16	72.86
.0161782	6	0.47	73.33
.0163558	6	0.47	73.80
.0164732	2	0.16	73.96
.0164784	1	0.08	74.04
.0172803	6	0.47	74.51
.0173346	1	0.08	74.59
.0173933	9	0.71	75.30
.0174927	9	0.71	76.00
.0176708	9	0.71	76.71
.0177397	5	0.39	77.10
.0181726	6	0.47	77.58
.0190019	11	0.87	78.44
.0194346	8	0.63	79.07
.0194439	1	0.08	79.15
.0203905	3	0.24	79.39
.0208214	10	0.79	80.17
.0211058	7	0.55	80.72
.0217466	1	0.08	80.80
.0222638	13	1.02	81.83
.0222946	17	1.34	83.16
.0225491	6	0.47	83.63
.0227317	12	0.94	84.58
.0229383	10	0.79	85.37
.0231102	15	1.18	86.55
.0231135	23	1.81	88.36
.0233705	31	2.44	90.79
.0234546	23	1.81	92.60
.0234876	26	2.05	94.65
.023789	19	1.49	96.14
.0242385	14	1.10	97.25
.0244431	12	0.94	98.19
.0245299	15	1.18	99.37
.0370642	2	0.16	99.53
.0494498	6	0.47	100.00

Table 15 - Tabulation of AVG_CASES

Asthma	Freq.	Percent	Cum.
0 = No	1,245	97.95	97.95
1 = Yes	26	2.05	100.00

Table 16 – Tabulation of P26_1

Lung Disease	Freq.	Percent	Cum.
0 = No	1,249	98.27	98.27
1 = Yes	22	1.73	100.00

Table 17 – Tabulation of P26_2

Heart Attack	Freq.	Percent	Cum.
0 = No	1,261	99.21	99.21
1 = Yes	10	0.79	100.00

Table 18 – Tabulation of P26_3

Heart Disease	Freq.	Percent	Cum.
0 = No	1,253	98.58	98.58
1 = Yes	18	1.42	100.00

Table 19 – Tabulation of P26_4

Hypertension	Freq.	Percent	Cum.
0 = No	1,154	90.79	90.79
1 = Yes	117	9.21	100.00

Table 20 – Tabulation of P26_5

Stroke	Freq.	Percent	Cum.
0 = No	1,261	99.21	99.21
1 = Yes	10	0.79	100.00

Table 21 – Tabulation of P26_6

Arthritis	Freq.	Percent	Cum.
0 = No	1,230	96.77	96.77
1 = Yes	41	3.23	100.00

Table 22 – Tabulation of P26_7

Back Pain	Freq.	Percent	Cum.
0 = No	1,239	97.48	97.48
1 = Yes	32	2.52	100.00

Table 23 – Tabulation of P26_8

Neck Pain	Freq.	Percent	Cum.
0 = No	1,255	98.74	98.74
1 = Yes	16	1.26	100.00

Table 24 – Tabulation of P26_9

Diabetes	Freq.	Percent	Cum.
0 = No	1,171	92.13	92.13
1 = Yes	100	7.87	100.00

Table 25 – Tabulation of P26_10

Allergies	Freq.	Percent	Cum.
0 = No	1,258	98.98	98.98
1 = Yes	13	1.02	100.00

Table 26 – Tabulation of P26_11

Hepatic Cirrhosis	Freq.	Percent	Cum.
0 = No	1,264	99.45	99.45
1 = Yes	7	0.55	100.00

Table 27 – Tabulation of P26_12

Bladder Control	Freq.	Percent	Cum.
0 = No	1,266	99.61	99.61
1 = Yes	5	0.39	100.00

Table 28 – Tabulation of P26_13

Kidney Disease	Freq.	Percent	Cum.
0 = No	1,260	99.13	99.13
1 = Yes	11	0.87	100.00

Table 29 – Tabulation of P26_14

Depression	Freq.	Percent	Cum.
0 = No	1,237	97.32	97.32
1 = Yes	34	2.68	100.00

Table 30 – Tabulation of P26_15

Cancer	Freq.	Percent	Cum.
0 = No	1,259	99.06	99.06
1 = Yes	12	0.94	100.00

Table 31 – Tabulation of P26_16

High Cholesterol	Freq.	Percent	Cum.
0 = No	1,268	99.76	99.76
1 = Yes	3	0.24	100.00

Table 32 – Tabulation of P26_17

Handwashing	Freq.	Percent	Cum.
0 = No	65	5.11	5.11
1 = Yes	1,206	94.89	100.00

Table 33 – Tabulation of CV1_1

Mask	Freq.	Percent	Cum.
0 = No	198	15.58	15.58
1 = Yes	1,073	84.42	100.00

Table 34 – Tabulation of CV1_2

No Handshaking	Freq.	Percent	Cum.
0 = No	212	16.68	16.68
1 = Yes	1,059	83.32	100.00

Table 35 – Tabulation of CV1_3

Social Distancing	Freq.	Percent	Cum.
0 = No	158	12.43	12.43
1 = Yes	1,113	87.57	100.00

Table 36 – Tabulation of CV1_4

No Crowds	Freq.	Percent	Cum.
0 = No	266	20.93	20.93
1 = Yes	1,005	79.07	100.00

Table 37 – Tabulation of CV1_5

No Supermarkets	Freq.	Percent	Cum.
0 = No	416	32.73	32.73
1 = Yes	855	67.27	100.00

Table 38 – Tabulation of CV1_6

No Friends	Freq.	Percent	Cum.
0 = No	507	39.89	39.89
1 = Yes	764	60.11	100.00

Table 39 – Tabulation of CV1_7

No Family	Freq.	Percent	Cum.
0 = No	569	44.77	44.77
1 = Yes	702	55.23	100.00

Table 40 – Tabulation of CV1_8

No Outside	Freq.	Percent	Cum.
0 = No	795	62.55	62.55
1 = Yes	476	37.45	100.00

Table 41 – Tabulation of CV1_9

No Behavior	Freq.	Percent	Cum.
0 = No	1,261	99.21	99.21
1 = Yes	10	0.79	100.00

Table 42 – Tabulation of CV1_10

General Risk	Freq.	Percent	Cum.
1 = No Risk	58	4.74	4.74
2 = Low Risk	239	19.53	24.26
3 – Risk Neutral	442	36.11	60.38
4 = High Risk	299	24.43	84.80
5 = Very High Risk	186	15.20	100.00

Table 43 – Tabulation of CV21

Individual Risk	Freq.	Percent	Cum.
1 = No Risk	61	4.97	4.97
2 = Low Risk	234	19.06	24.02
3 – Risk Neutral	418	34.04	58.06
4 = High Risk	300	24.43	82.49
5 = Very High Risk	215	17.51	100.00

Table 44 – Tabulation of CV22

Family Risk	Freq.	Percent	Cum.
1 = No Risk	25	2.04	2.04
2 = Low Risk	190	15.54	17.58
3 – Risk Neutral	439	35.90	53.48
4 = High Risk	341	27.88	81.36
5 = Very High Risk	228	18.64	100.00

Table 45 – Tabulation of CV23

Community Risk	Freq.	Percent	Cum.
1 = No Risk	21	1.75	1.75
2 = Low Risk	170	14.19	15.94
3 = Risk Neutral	439	36.64	52.59
4 = High Risk	381	31.80	84.39
5 = Very High Risk	187	15.61	100.00

Table 46 – Tabulation of CV24

Private Health Insurance	Freq.	Percent	Cum.
0 = Yes	189	14.91	14.91
1 = No	1,079	85.09	100.00

Table 47 – Tabulation of P29

Smoke	Freq.	Percent	Cum.
0 = Smoke Daily	394	31.55	31.55
1 = Smoke Occasionally	33	2.64	34.19
2 = Have smoked, but currently does not smoke	116	9.29	43.47
3 = Never Smoked	706	56.53	100.00

Table 48 – Tabulation of P32

Drink	Freq.	Percent	Cum.
0 = Every day or almost every day	306	24.64	24.64
1 = Between 5 to 6 days a week	48	3.86	28.50
2 = Between 3 to 4 days a week	106	8.53	37.04
3 = Between 2 to 1 days a week	126	10.14	47.18
4 = Between 2 to 3 times a month	135	10.87	58.05
5 = Once a month	43	3.46	61.51
6 = Less than once a month	153	12.32	73.83
7 = Has not consumed in the last 12 months, due to having stopped consuming alcohol	59	4.75	78.58
8 = Never consumer, or just occasionally tasting it	266	21.42	100.00

Table 49 – Tabulation of P33

General Health State	Freq.	Percent	Cum.
0 = Very good	221	17.39	17.39
1 = Good	648	50.98	68.37
2 = Reasonable	340	26.75	95.12
3 = Bad	59	4.64	99.76
4 = Very bad	3	0.24	100.00

Table 50 – Tabulation of P34

Gender	Freq.	Percent	Cum.
0 = Male	602	47.36	47.36
1 = Female	669	52.64	100.00

Table 51 – Tabulation of D1_gender

Age by Interval	Freq.	Percent	Cum.
0 = Ages between 15 and 24 years old	188	14.79	14.79
1 = Ages between 25 and 34 years old	203	15.97	30.76
2 = Ages between 35 and 44 years old	249	19.59	50.35
3 = Ages between 45 and 54 years old	226	17.78	68.14
4 = Ages between 55 and 64 years old	217	17.07	85.21
5 = Ages between 65 and 74 years old	103	8.10	93.31
6 = More than 75 years old	85	6.69	100.00

Table 52 – Tabulation of D2_age_brackets

Household Members by Interval	Freq.	Percent	Cum.
0 = Up to 2 members	608	47.84	47.84
1 = Between 3 and 4 members	566	44.53	92.37
2 = 5 or more members	97	7.63	100.00

Table 53 – Tabulation of D3_brackets

Education	Freq.	Percent	Cum.
0 = University/ Post-Graduate/ Master's/ Doctorate	118	9.28	9.28
1 = Middle Course/ Polytechnic Degree	15	1.18	10.46
2 = Participation in Higher Education/ Middle	26	2.05	12.51
3 = 12th Grade (7th Year of High School/ 11th Grade)	375	29.50	42.01
4 = 9th Grade (5th Year of High School)	280	22.03	64.04
5 = 6th Grade (2nd Year of High School)	164	12.90	76.95
6 = Complete Primary Education	244	19.20	96.14
7 = Incomplete Primary Education/ Analphabet	49	3.86	100.00

Table 54 – Tabulation of D5

Occupation	Freq.	Percent	Cum.
0 = Self-Employed	132	10.39	10.39
1 = Employee	598	47.05	57.44
2 = Unemployed	150	11.80	69.24
3 = Retired	245	19.28	88.51
4 = Stay at Home/ Responsible only for dealing with household issues	49	3.86	92.37
5 = Student	97	7.63	100.00

Table 55 – Tabulation of D6

Income	Freq.	Percent	Cum.
0 = Lesser or equal to 450 euros	52	8.05	8.05
1 = 451 - 650 euros	96	14.86	22.91
2 = 651 - 800 euros	107	16.56	39.47
3 = 801 – 1,000 euros	83	12.85	52.32
4 = 1,001 – 1,200 euros	115	17.80	70.12
5 = 1,201 – 1,400 euros	81	12.54	82.66
6 = 1,401 – 1,700 euros	54	8.36	91.02
7 = 1,701 – 2,000 euros	36	5.57	96.59
8 = 2,001 – 2,700 euros	15	2.32	98.92
9 = More than 2,700 euros	7	1.08	100.00

Table 56 – Tabulation of D7

Region	Freq.	Percent	Cum.
0 = Northern Seaside	265	20.85	20.85
1 = Great Oporto	156	12.27	33.12
2 = Interior	168	13.22	46.34
3 = Central Seaside	219	17.23	63.57
4 = Great Lisbon	360	28.32	91.90
5 = Alentejo	52	4.09	95.99
6 = Algarve	51	4.01	100.00

Table 57 – Tabulation of region

GFK2020	Number of Cases Data
Northern Seaside	Regional Health Authority of the North
Great Oporto	Regional Health Authority of the North
Interior	Regional Health Authority of the North
Central Seaside	Regional Health Authority of the Centre
Great Lisbon	Regional Health Authority of Lisbon and the Tagus Valley
Alentejo	Regional Health Authority of the Alentejo
Algarve	Regional Health Authority of the Algarve

Table 58 –Match between the regions registered in the GFK2020 and those used by the Portuguese National Health Authorities to report new cases (downloaded from https://github.com/dssg-pt/covid19pt-data/blob/master/data.csv)

Annex III – Equations

$$AVG\_CASES_i = \frac{\sum_{z=1}^{7} \frac{Cases_{j(n-z)} - Cases_{j(n-z-1)}}{Cases_{j(n-z-1)}}}{7}, \quad j = region\ of\ individual\ i, \quad n = the\ day\ the\ survey\ was\ taken\ by\ individual\ i$$

Equation 1 – Formula for the variable AVG_CASES

$$\begin{aligned} Individual\ Risk\ Binary\ Form_i &= \beta_0 + \beta_1 AVG\_CASES_i + \beta_2 Asthma_i + \beta_3 Lung\ Disease_i \\ &+ \beta_4 Heart\ Attack_i + \beta_5 Heart\ Disease_i + \beta_6 Hypertension_i \\ &+ \beta_7 Stroke_i + \beta_8 Arthritis_i + \beta_9 Back\ Pain_i + \beta_{10} Neck\ Pain_i \\ &+ \beta_{11} Diabetes_i + \beta_{12} Allergies_i + \beta_{13} Hepatic\ Cirrhosis_i \\ &+ \beta_{14} Bladder\ Control_i + \beta_{15} Kidney\ Disease_i + \beta_{16} Depression_i \\ &+ \beta_{17} Cancer_i + \beta_{18} High\ Cholesterol_i \\ &+ \beta_{19} Private\ Health\ Insurance_i + \beta_{20} i.Smoke_i + \beta_{21} i.Drink_i \\ &+ \beta_{22} General\ Health\ State_i + \beta_{23} Gender_i + \beta_{24} i.Age\ by\ Intervals \\ &+ \beta_{25} i.Household\ Members\ by\ Interval_i + \beta_{26} i.Education_i \\ &+ \beta_{27} i.Occupation_i + \beta_{28} i.Region_i + \epsilon_i \end{aligned}$$

Equation 2 – Formula for the regression of the binary forms of Individual Risk on all covariates through LPM, SUR and MLE

$$
\begin{aligned}
&Family\ Risk\ Binary\ Form_i \\
&\qquad = \beta_0 + \beta_1 AVG\_CASES_i + \beta_2 Asthma_i + \beta_3 Lung\ Disease_i \\
&\qquad + \beta_4 Heart\ Attack_i + \beta_5 Heart\ Disease_i + \beta_6 Hypertension_i \\
&\qquad + \beta_7 Stroke_i + \beta_8 Arthritis_i + \beta_9 Back\ Pain_i + \beta_{10} Neck\ Pain_i \\
&\qquad + \beta_{11} Diabetes_i + \beta_{12} Allergies_i + \beta_{13} Hepatic\ Cirrhosis_i \\
&\qquad + \beta_{14} Bladder\ Control_i + \beta_{15} Kidney\ Disease_i + \beta_{16} Depression_i \\
&\qquad + \beta_{17} Cancer_i + \beta_{18} High\ Cholesterol_i \\
&\qquad + \beta_{19} Private\ Health\ Insurance_i + \beta_{20} i.Smoke_i + \beta_{21} i.Drink_i \\
&\qquad + \beta_{22} General\ Health\ State_i + \beta_{23} Gender_i + \beta_{24} i.Age\ by\ Intervals \\
&\qquad + \beta_{25} i.Household\ Members\ by\ Interval_i + \beta_{26} i.Education_i \\
&\qquad + \beta_{27} i.Occupation_i + \beta_{28} i.Region_i + \epsilon_i
\end{aligned}
$$

Equation 3 – Formula for the regression of the binary forms of Family Risk on all covariates through LPM, SUR and MLE

$$
\begin{aligned}
Community\ Risk\ Binary\ Form_i \\
&= \beta_0 + \beta_1 AVG\_CASES_i + \beta_2 Asthma_i + \beta_3 Lung\ Disease_i \\
&+ \beta_4 Heart\ Attack_i + \beta_5 Heart\ Disease_i + \beta_6 Hypertension_i \\
&+ \beta_7 Stroke_i + \beta_8 Arthritis_i + \beta_9 Back\ Pain_i + \beta_{10} Neck\ Pain_i \\
&+ \beta_{11} Diabetes_i + \beta_{12} Allergies_i + \beta_{13} Hepatic\ Cirrhosis_i \\
&+ \beta_{14} Bladder\ Control_i + \beta_{15} Kidney\ Disease_i + \beta_{16} Depression_i \\
&+ \beta_{17} Cancer_i + \beta_{18} High\ Cholesterol_i \\
&+ \beta_{19} Private\ Health\ Insurance_i + \beta_{20} i.Smoke_i + \beta_{21} i.Drink_i \\
&+ \beta_{22} General\ Health\ State_i + \beta_{23} Gender_i + \beta_{24} i.Age\ by\ Intervals \\
&+ \beta_{25} i.Household\ Members\ by\ Interval_i + \beta_{26} i.Education_i \\
&+ \beta_{27} i.Occupation_i + \beta_{28} i.Region_i + \epsilon_i
\end{aligned}
$$

Equation 4 – Formula for the regression of Individual Risk on all covariates through LPM, SUR and MLE

Annex IV – Interpretation of Estimated Coefficients

All three notions of risk perception, considering the different levels of proximity, give the following average (at mean values for the MLE), *ceteris paribus* analysis:

Individual Risk

First, at the mean values of all its covariates, the predicted probability that the binary form of *Individual Risk* is 1 is 43.8% through LPM and SUR, and 44.8% through MLE (*p-values* of 0.000),[19] meaning that the average individual has a higher probability of perceiving a high risk regarding its personal health from COVID-19. Furthermore, the LPM and SUR present an *adjusted R-squared* of 0.248 and a *R-squared* of 0.285, while the MLE presents an *adjusted R-squared* of 0.155 and a *Pseudo R-squared* of 0.239. The regressed models might, however, have included too many unnecessary and dummies variables. Nonetheless, given the literary references that pointed towards the possible importance of the included variables, it was deemed relevant to include all the presented variables under the current format so as to create a more comprehensive profile for risk perceptions.

Average Daily Percentage Change in the Number of COVID Cases

An increase in the average daily percentage change in the number of COVID cases in the week immediately before a given person took the survey was associated with an increase in the in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health (positive coefficient) according to all three models. Nonetheless, none of the estimations is statistically significant at the usual confidence levels.

[19] Marginal Analysis in Table 13 in Appendix I

Chronic Conditions

Under the LPM and SUR models, most chronic conditions are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health (positive coefficient). Exceptions include Stroke, Back Pain, Hepatic Cirrhosis, Bladder Control, Depression and High Cholesterol. Of all pre-existing conditions only Asthma, Diabetes and Allergies are statistically significant at a minimum 90% confidence level (the first only according the MLE and the rest considering all three models). It is interesting, however, to see how some diseases related to the organs most affected by COVID-19 (lungs and heart) and to mental health are associated with a decrease in the probability of one having a high individual risk perception.

Private Health Insurance

According to LPM, SUR and MLE, having a private insurance scheme is associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health, although they are not statistically significant (*p-value* of 0.860, 0.857 and 0.764, respectively). Here, it could be the case that those acquiring private health insurance are also those most worried with the pandemic crisis, which might represent some kind of adverse selection situation even if in a non-observable manner.

Smoke

In general, people reporting to smoke more frequently are associated with an increase in the probability that they have a high perception of the risk COVID-19 poses to their individual health in all three models. Moreover, only the category associated with smoking occasionally is statistically significant (with *p*-values of 0.045 through the LPM, 0.040 through the SUR and 0.032 through the MLE), and it is interesting to point out that the less frequently a person reports to smoke, the lower is the size of the associated increase, meaning that one's tendency to not smoke seems to be positively associated with their tendency of perceiving a low risk-level regarding COVID-19, which was expectable as smoking affects the organs most harmed by the new Coronavirus, meaning that it presents a threat to one's individual life.

Drink

According to LPM, SUR and MLE, only people reporting to drink 5 to 6 days a week or to never having had a drink are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health, while all other alcohol consumption frequencies are associated with a decrease in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health (negative coefficient). This information translates into a parabolic shape that interestingly joins those that consume the most and the least in the same group in terms of the sign of the effect on the likelihood of perceiving a high risk-level concerning the virus. However, only the coefficient for individuals reporting to drink once or twice a week are statistically significant considering a 95% level of confidence (*p-values* of 0.017, 0.015 and 0.013, respectively).

General Health State

Regarding one's reported general healthcare state, any kind of notion is associated with a positive coefficient, with those reporting worse health states being associated to higher increases, on average, *ceteris paribus*, in all three approaches. Another interesting point is that reporting either a reasonable or a bad general health state are statistically significant at a 99% confidence level, while all other states are non-significant at the usual confidence levels.

Gender

According to LPM, SUR and MLE, females are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health (positive coefficient), with no statistically significant estimates.

Age by Intervals

In terms of age differences, all age levels are associated to an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health through the LPM, SUR and MLE. Furthermore, all age intervals are significant at a minimum 90% confidence level. It is interesting to see how the elderly (65 and older) are associated with a much higher increase than younger intervals. Unlike previous literature, that posited that the elderly could be more prone to perceiving a lower risk level due to some kind of coping mechanism through which they focus on the bright side of things while they try to maximize their psychological well-being during this incredibly stressful period, data seems to point towards what would be expectable *à priori*: that older individuals, those most vulnerable to the negative consequences of the virus, are more likely to be aware of the vulnerability and perceive a higher risk-level.

Household Members by Intervals

Through the LPM, SUR and MLE, households with five or more members are associated with a decrease in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health, while households with three or four members are associated with a positive coefficient, although with no statistically significant estimates. This result is still worth pointing out since one could think that living with more people could put them in more danger in terms of being infected with the new Coronavirus due to the possible necessity of more people to contact with elements from outside that household. Nonetheless, this effect seems to be ambiguous, given by the different signs of the coefficients of both sizes, on whether that is in fact the case or if these elements believe that they guarantee the conditions for a group immunity within their family, that they are a constrict group with a low level of external contacts, which could make them less susceptible to perceiving a high risk related with the virus.

Education

Individuals reporting higher levels of education are associated with a higher positive coefficient while those reporting lower levels of education are associated with a lower increase in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health (positive coefficient). This might be an indication that lower levels of education use more of a natural intuition for risk and, thus, might be less susceptible to believing and reacting to external information regarding the evolution of the pandemic crisis that could eventually make them more prone to perceiving a high risk to their individual health. On the other side of the spectrum, this could also be a sign that people with higher education levels are more aware or have better tools to dissect all that external information about the virus, thus being more likely to worry about it. Finally, only the estimates regarding individuals reporting to having completed a Middle Course/ Polytechnic Degree are statistically significant (*p-values* of 0.054 through the LPM, 0.048 through the SUR and 0.044 through the MLE).

Occupation

According to LPM and SUR, any of the occupational settings presented in the survey, except for being unemployed, is associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health (positive coefficient), while according to the MLE both being employed and unemployed is associated with negative coefficients. However, none of these estimates are statistically significant.

Region

Taking into account differences in Portuguese regions, there is no specific region associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their individual health (i.e. all present negative coefficients). Only the Great Lisbon region and Alentejo's coefficients are not statistically significant for a minimum confidence interval of 90% in all three models. However, it might either be the case that, due to the size of the pandemic crisis, one's region is not perceived as being associated with a particularly high chance of suffering due to the virus, at least at a personal level, or that the familiarity of one's region of residence leads them to be less prone to perceiving a high risk associated with the virus.

Family Risk

Concerning *Family Risk* in binary form, at the mean values of all its covariates, the predicted probability of it being 1 is 42.5% through LPM and SUR, and 41.9% through MLE (*p-values* of 0.000),[20] i.e. that the average individual has a higher probability of perceiving a high risk regarding its personal health from COVID-19. Furthermore, the LPM and SUR present an *adjusted R-squared* of 0.222 and a *R-squared* of 0.261, while the MLE presents an *adjusted R-squared* of 0.127 and a *Pseudo R-squared* of 0.211. Once again, there might have been too many variables in dummy format included. Nonetheless, it was deemed that there were benefits in including them so as to create the more comprehensive profile for risk perceptions possible.

Average Daily Percentage Change in the Number of COVID Cases

Similar to the case of *Individual Risk*, an increase in the average daily percentage change in the number of COVID cases in the week immediately before a given person took the

[20] Marginal Analysis in Table 13 in Appendix I

survey is associated with an increase in the probability that a given individual has a high perception of the risk COVID-19 poses to their family's health (positive coefficient) according to all three models, again with no estimation being statistically significant.

Chronic Conditions

When thinking of the risk COVID-19 poses to one's family's health, most chronic conditions are associated with an increase in the probability that they have a high perception (positive coefficient). The exceptions are now reporting to have or have had a Stroke, Arthritis, Neck Pain, Hepatic Cirrhosis and Depression. Of all pre-existing conditions, once again, only Diabetes and Allergies are statistically significant for all three models (at a minimum 95% confidence level), and Neck Pain is statistically significant according to the MLE (*p-value* of 0.092).

Private Health Insurance

Having a private insurance scheme is associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health (positive coefficient), although it is not statistically significant (*p-value* of 0.901 through LPM, 0.844 through MLE and 0.898 through SUR). As in the *Individual Risk*, this phenomenon could be a reflection of some case of self-selection of those more preoccupied with the negative effects of the pandemic crisis to acquire a private health insurance.

Smoke

According to the LPM and SUR, all smoking frequencies are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health (positive coefficient). However, according to the MLE, only those reporting to never having smoked in their lives are associated with a negative coefficient. Although none of these estimates are statistically significant at any level of smoking frequency, it is interesting to point out the decrease in the magnitude of the effect as one reports to smoke less frequently, similar to the results presented regarding the individual risk notion but with a slightly less steeper decrease. Thus, these results seem to once again behave as one would expect *à priori*, with the probability that they have a high risk perception regarding COVID-19 being compatible with the idea that if someone smokes they should be more vulnerable to the disease and, thus, have a relatively higher risk perception.

Drink

As another behaviour usually associated with severe health problems, on the one hand, reporting to drink 5 to 6 days a week or 2 to 3 times a month is associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health (positive coefficient). On the other hand, all other alcohol consumption frequencies are associated with a decrease in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health (negative coefficient). Moreover, these effects depict a somewhat parabolic behaviour as the drinking frequency decreases, while slightly less pronounced that that in the individual risk notion, and none of these estimates are statistically significant considering the standard levels of confidence (90%, 95% or 99%).

General Health State

Only the coefficient regarding those reporting to have a generally good health state are associated with a decrease in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health (negative coefficient), with those reporting worse health states being associated to higher increases, on average, *ceteris paribus*. The coefficients of a Good and Very Bad Health states are not statistically significant through any of the models, while the remaining states are relevant at a maximum *p-value* of 0.010.

Gender

According to all three models, females are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health (positive coefficient), with a statistically significant estimate at a 90% confidence interval according to the LPM and SUR (*p-values* of 0.063 and 0.057, respectively), and at a 95% confidence level through the MLE (*p-value* of 0.040).

Age by Intervals

Considering the different age brackets used in the regression process, all ages are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health, with all estimates being statistically significant at a minimum confidence level of 90%. Furthermore, the magnitude of the coefficients is higher in intervals of older ages. Here, data seems to contradict existing literature like previously observed for *Individual Risk*.

Household Members by Intervals

Keeping in mind the number of reported household members, the intervals above 2 are associated with a negative coefficient and larger households are associated with a greater decrease than smaller households, which could support the group immunity effect previously exposed. However, these estimates are not statistically significant.

Education

In terms of the different levels of education, higher levels of education are associated with a higher increase in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health, while those reporting lower levels of education are associated with a relatively smaller increase. In other words, more educated people might be more susceptible to being more worried with COVID-19's impact on their family. Moreover, only the estimates for holding a Middle Course/ Polytechnic Degree, having completed the 9th Grade and having completed Primary Education are statistically significant according to all three models (the first at a 90% confidence level and the remaining at a 95% confidence interval).

Occupation

According to LPM and SUR, only reporting to being retired is associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health (positive coefficient) while, according to the MLE, all coefficients are negative. However, only being employed or unemployed are statistically significant through all models at a 95% Confidence level.

Region

Regarding the geographical distribution of the respondents, all regions but the Great Lisbon area and Alentejo are associated with a decrease in the probability that a given person has a high perception of the risk COVID-19 poses to their family's health according to all models. Furthermore, most regions' estimates are statistically insignificant except the Interior through the MLE (*p-value* of 0.098) and the Algarve region through all models (at a 99% confidence level).

Community Risk

Finally, and at the farthest level of proximity, considering the mean values of all its covariates, the predicted probability for the binary form of *Community Risk* being 1 is 51.3% through LPM and SUR, and 51.3% through MLE (*p-values* of 0.000).[21] This

[21] Marginal Analysis in Table 8 in Appendix I

means that the average individual has a higher probability of perceiving a high risk regarding their community's health from COVID-19. Furthermore, the LPM and SUR present an *adjusted R-squared* of 0.108 and a *R-squared* of 0.153, while the MLE presents an *adjusted R-squared* of 0.046 and a *Pseudo R-squared* of 0.128. Because of the aforementioned literary references, it was deemed relevant to include all the presented variables, in their presented form, in order to create the best possible profile for risk perceptions, even with it implying that the values of the *adjusted R-squares* would be small.

Average Daily Percentage Change in the Number of COVID Cases

Unlike in other risk perceptions, an increase in the average daily percentage change in the number of COVID cases in the week immediately before a given person took the survey is associated with a decrease in the in the probability that they have a high perception of the risk COVID-19 poses to their community's health (negative coefficient) according to all three models, with no estimate being statistically significant.

Chronic Conditions

According to all models, most chronic conditions are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their community's health (positive coefficient), with the exception of Lung Disease, Heart Attack, Heart Disease (only through the MLE), Arthritis, Neck Pain, Hepatic Cirrhosis, Diabetes and Cancer. Of all pre-existing conditions only Diabetes (at a 99% confidence level), Allergies (at a 90% confidence level through the LPM and MLE), Bladder Control (at a 95% confidence level through the LPM) and Kidney Disease (at a 90% confidence level through the LPM) are statistically significant.

Private Health Insurance

Reporting to have an health insurance scheme is associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their community's health (positive coefficient), although they are not statistically significant (*p-value* of 0.466, 0.453 and 0.470, through LPM, SUR and MLE respectively). This is one more result in favour of the suspicions of some type of adverse selection or self-selection being the reason for this association.

Smoke

In general, smoking occasionally is associated with a positive coefficient while the other levels of smoking frequency are associated with a negative sign. Although none of these estimates are statistically significant at any level of smoking frequency, it makes a similar point in agreeing with *à* priori expectations in like the other levels of proximity.

Drink

Regarding alcohol consumption, drinking either 5 to 6 or 1 to 2 days a week is associated with a decrease in the probability that a given person has a high perception of the risk COVID-19 poses to their community's health (negative coefficient), while individuals drinking at any of the remaining frequencies are associated with an increase the probability that a given person has a high perception of the risk COVID-19 poses to their community's health (positive coefficient). However, only the estimate for drinking between 1 and 2 days per week is statistically significant at a 95% confidence level (*p-values* of 0.017 through the LPM, 0.015 through the SUR and 0.013 through the MLE). At this proximity level, there does not seem to be a parabolic tendency like before, and lower levels of alcoholic consumption seem to be associated with a higher chance of being more with the new Coronavirus.

General Health State

Concerning one's notion of their general health state, all possible views on this perception are associated with a positive coefficient. However, the general tendency of increase in the magnitude of the increment with the worsening of the perceived health state seems to be less persistent than before, disappearing for those reporting a Very Bad general health state. Another interesting point is that, once again, only the coefficients of reporting a Reasonable or Bad general heath state are statistically significant to all three models, at a 99% confidence level.

Gender

According to LPM and SUR, females are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their community's health (positive coefficient). Nonetheless, none of the estimates is statistically significant.

Age by Intervals

In terms of age differences, all age levels are associated with an increase in the probability that a given person has a high perception of the risk COVID-19 poses to their community's health (positive coefficient). Furthermore, all estimates are statistically significant at a minimum confidence level of 90%, and the magnitude of the coefficient is relatively higher in the interval of ages above 65 years than at other age brackets, which supports the reasoning passed on by the other risk notions.

Household Members by Intervals

Unlike at other proximity levels, families with three or four elements are associated with a decrease in the probability that a given person has a high perception of the risk COVID-19 poses to their community health according to all models, but the opposite happens for families with 5 or more members. Although none of the estimates is statistically significant, it is worth pointing out the fact that these results completely invest those previously observed.

Education

Again, unlike in the other risk notions, and regarding the impact of one's level of education, higher levels are associated with a lower positive coefficient than lower levels of education. The magnitude of this associating seems to increase further as people report to have studied less than until the 6th grade. Only the coefficients of reporting to not have completed any educational level (at a 99% confidence level) and of having completed the 12th Grade (at a 95% confidence level) are statistically significant.

Occupation

Analysing the impact of one's occupation on *Community Risk*, on the one hand, being employed or retired is associated with a decrease in the probability that a given person has a high perception of the risk COVID-19 poses to their community's health (negative coefficient). On the other hand, all other occupations are associated with a positive coefficient. However, none of these estimates is statistically significant.

Region

Regarding one's region, only the Interior, the Central Seaside and Algarve are not associated with a positive coefficient. Furthermore, only the Central Seaside, Alentejo and Algarve are statistically significant for a minimum confidence interval of 95%.